EXPANDING QUANTITATIVE TOOLS FOR SECONDARY STRUCTURE ANALYSIS OF DNA APTAMER CANDIDATES SELECTED VIA COMPELS

A Dissertation
Presented to
The Academic Faculty

by

Richard Stanley Sullivan III

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Materials Science and Engineering

Georgia Institute of Technology

December, 2017

EXPANDING QUANTITATIVE TOOLS FOR SECONDARY STRUCTURE ANALYSIS OF DNA APTAMER CANDIDATES SELECTED VIA COMPELS

Approved by:

Dr. Valeria Milam, Advisor
School of Materials Science and Engineering,
Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Zhiqun Lin
School of Materials Science and Engineering
Georgia Institute of Technology

Dr. Mostafa El-Sayed
School of Chemistry
Georgia Institute of Technology

Dr. Christopher Muhlstein
School of Materials Science and Engineering
Georgia Institute of Technology

Dr. Dong Qin
School of Materials Science and Engineering
Georgia Institute of Technology

Dr. Rajesh Naik
Air Force Research Laboratory,
Materials & Manufacturing Directorate,
Soft Matter Materials Branch
Wright Patterson AFB

Date Approved: August 14, 2017

“The essence of science lies not in discovering facts, but in discovering new ways of thinking about them.”

- W.L. Bragg

“... [A] knowledge of sequences could contribute much to our understanding of living matter.”

-Frederick Sanger

“Success consists of going from failure to failure without loss of enthusiasm.”

-Winston Churchill

ACKNOWLEDGEMENTS

I would like to thank all my friends and family who have remained a constant support over the years I have pursued the life of a scientist, researcher, and engineer with all its inherent struggles. In particular, I would like to thank Dr. Valeria Milam for her mentorship, critical thinking, and constant belief in me, even when things didn't go as planned; Dr. Maeling Murphy for her philosophical and intellectual and technical support and expertise over the course of our related studies; and Mary-Catherine Adams for her assistance in processing data, moral support in "ticking clock" domain of thesis writing, and without whom much of this would not have been possible; and Julia and Richard Sullivan, whose love, parental guidance, support, friendship, and instruction that long instilled in me a fondness of learning and philosophy, have molded me into the man I am today. I would also like to thank both Dr. Bryan Baker and Dr. James Hardin for their intellectually stimulating conversation and assistance in initial comprehension mechanics of DNA at the start of my graduate career. I would like to thank my thesis committee members for their valuable time in reviewing and providing feedback on my work. I would like to thank my undergraduate students Alyssa Bronson and Jonathan Peraza for their help in implementing details of these analyses. Lastly, I would like to thank Audra Barton for her much needed love and support.

I would like to gratefully acknowledge the financial support I have received from the Department of Materials Science and Engineering, Air-Force Office of Scientific Research (AFSOR), and the Center for Pediatric Nanomedicine that enabled me to finish my research here at Georgia Institute of Technology. I would like to thank Dr. Naresh

Thadahani, Susan Bowman, Teresa Nelson, and Mechelle Kitchings for their assistance and support over my time here at Georgia Tech. I would also like to thank Dr. Rajesh Naik and his group at the Air Force research labs, especially Dr. Patrick Dennis and Sharon Jones, for hosting my research studies at Wright-Patterson AFB in 2012 and for their continual input throughout my PhD studies. Additionally, I would like to acknowledge the excellent Core Lab facilities and staff of the Institute of Bioengineering and Biosciences (IBB) which enabled my many flow cytometry studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DNA	deoxyribonucleic acid
RNA	ribonucleic acid
LNA	locked nucleic acid
XNA	xeno nucleic acid
ssDNA	single-stranded deoxyribonucleic acid
dsDNA	double-stranded deoxyribonucleic acid
SELEX	Systematic Evolution of Ligands by Exponential enrichment
PCR	polymerase chain reaction
RT-PCR	reverse transcription polymerase chain reaction
VEGF	vascular endothelial growth factor
ePCR	emulsion polymerase chain reaction
ddPCR	digital droplet polymerase chain reaction
A	adenine
T	thymine
C	cytosine
G	guanine
dNTP	deoxynucleotide triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
NTP	nucleoside triphosphate

ATP	adenosine triphosphate
CTP	cytidine triphosphate
GTP	guanosine triphosphate
UTP	uridine triphosphate
dNMP	deoxynucleoside monophosphate
dAMP	deoxyadenosine monophosphate
dCMP	deoxycytidine monophosphate
dGMP	deoxyguanosine monophosphate
dTMP	deoxythymidine monophosphate
Au	gold
AuNP	gold nanoparticle
AuNR	gold nanorod
LE	Lambda Exonuclease
PAGE	polyacrylamide gel electrophoresis
CTAB	cetyl trimethylammonium bromide
TEM	transmission electron microscopy
UV-Vis	ultraviolet-visible spectroscopy
CompELS	Competition Enhanced Ligand Selection
NGS	Next Generation Sequencing
MSA	multiple sequence alignment
EM	expectation maximization
MEME	Multiple EM for Motif Elicitation
P-value	Poisson value
E-value	expectation value
PSSM	position specific score matrix

PWM	position weight matrix
SPR	surface plasmon resonance
SERS	surface enhanced raman scattering
FTIR	Fourier transform infrared spectroscopy
SSE	secondary structure element
SSF	secondary structure family
SS$	secondary structure string
ROS	reactive oxygen species
PTP	protein tyrosine phosphatases
MKP3	mitogen-activated protein kinase
$MKP3_{red}$	reduced MKP3
$MKP3_{ox}$	oxidized MKP3

LIST OF SYMBOLS

T_m	melting temperature
ΔG	Gibbs Free Energy
ΔH	enthalpy
ΔS	entropy
R	universal gas constant
T	absolute temperature
K_a	equilibrium association constant
K_d	equilibrium dissociation constant
r	Michaelis-Menten constant
z	Z-score
i	sequence number 1 to N out of set of sequences
N	number of sequences in a sequence set
I	index function
j	is the base number 1 to l in a sequence of length l
k	alphabet of possible SSE (S, H, L, I, G, D, and M)
$X_{i,j}$	the SSE equal to a value of k for sequence i at base number j
θ_j	psuedocounting or additive smoothing function at j
α	the pseudocount value
$\mu_{k,j}$	background incidence rate of an SSE equal to value k at base number j
$f_{k,j}$	incidence rate of an SSE equal to value k at base number j
Δ	$f_{k,j} - \mu_{k,j}$
$d_{n,n+1}$	average number of bases between secondary structure domain n and $n+1$

d_{prior}	average number of bases between a secondary structure domain and any previously occurring domain
d_{next}	average number of bases between a secondary structure domain and any next occurring secondary structure domain
$W_{k,j}$	position weight matrix (PWM)
Hj	normalization of the PWM to PSSM
$S_{k,j}$	position specific score matrix (PSSM)
τ_i	sequence specific score value from a PSSM

SUMMARY

The inclusion of DNA in materials systems has long evolved past its known biological function in cells and has garnered expanding interest as a macromolecular, recognition-based tool. To further the understanding of single-stranded DNA sequences called aptamers that bind with high affinity and specificity to non-nucleotide targets, this dissertation develops tools to identify, quantify, and classify secondary structure elements (SSE) and secondary structure families (SSF) that distinguish aptamer candidates from a large pool of random sequences. Chapter 1 provides an overview of oligonucleotide aptamers, the conventional method of their discovery, non-nucleotide target choices, and examples of aptamer-target binding characterization approaches. Chapter 2 investigates the effects of various nucleic acid additions during the seed mediated growth of gold nanorods (AuNR) on observed ultraviolet-visible (UV-Vis) spectra. Chapter 3 describes a non-evolutionary selection process we call Competition-Enhanced Ligand Selection (CompELS) for identifying DNA aptamers against gold nanorods (AuNR) and undertaking primary structure analysis of sequences identified. Chapter 4 develops analytical methods for secondary structure analysis of DNA aptamers selected through a non-evolutionary approach. Chapter 5 presents a non-evolutionary, two-stage CompELS approach to identify DNA aptamers against a protein target lacking antibody options due to important transient oxidation events during cell signaling events. Chapter 6 provides concluding comments and thoughts on potential further development and application of this work.

CHAPTER 1. INTRODUCTION: A LITERATURE REVIEW OF OLIGONUCELOTIDE APTAMERS

Deoxyribonucleic acids (DNAs) use as recognition-based detection and even as a therapeutic agents has only been intensifying since the publication of screening methods developed simultaneously and independently in the laboratories of G.F. Joyce, J.W. Szostak, and L. Gold almost three decades ago which enabled identification of oligonucleotides which could bind with high affinity and specificity to non-nucleotide targets. This pioneering work has since enabled the identification of oligonucleotides that can recognize targets which include ions, small molecules, proteins, and even whole cells. The interest in oligonucleotides persists due to their extremely unique and customizable properties that allow facile modification of both physical and chemical properties for tailoring of molecular interactions with both biological and synthetic materials systems. The following literature review will outline the distinctive properties of DNA and other nucleic acids and their evolved form, the aptamer. The conventional methods for identification of aptamers as developed by Szostak and Gold will be discussed along with common examples of aptamer targets in literature. Finally, current state-of-the-art characterization of aptamer-target binding will be reviewed.

1.1 Oligonucleotides and Their Properties

1.1.1 DNA and its Properties

1.1.1.1 DNA Chemical and Physical Structure and Properties

An oligonucleotide is a biomacromolecule or biopolymer chain with repeat units consisting of series of nucleotides. For DNA, each repeat unit or nucleotide consists of an alternating deoxyribose sugar and a negatively charged phosphate group, which comprise the backbone, and a purine or pyrimidine base side group which consist of either adenine (A) or guanine (G) for purine bases or cytosine (C) or thymine (T) for pyrimidine bases. Figure 1.1.1 shows the chemical structure of nucleotides for DNA. The size of each nucleotide is ~0.6 nm.

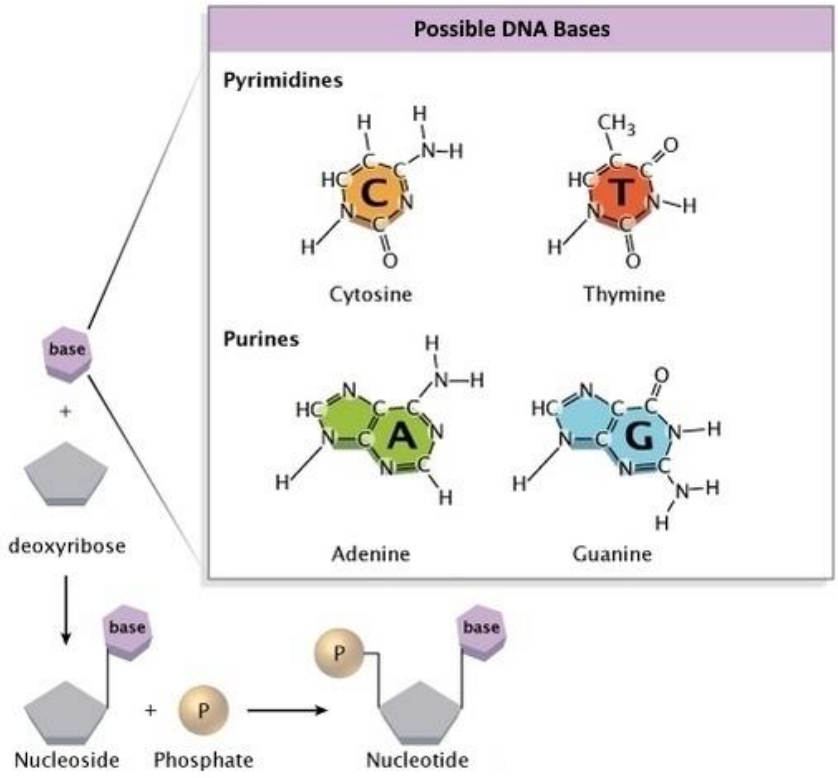


Figure 1.1.1. Illustration the chemical structure of nucleotides for DNA. Modified from [1]

The order in which nucleotides appear in the oligonucleotide, also known as the sequence or its primary structure, affects the physio-chemical properties of the oligonucleotide.[2]

In a living cellular context, a DNA sequence is the basis for the storage of genetic information, and is naturally found forming a helical structure or duplex with another oppositely oriented (5' to 3') and complementary DNA sequence. This iconic helical structure of DNA is made possible by hydrogen bonding between the bases in which specific complementarity between one purine base and a pyrimidine base is present and allows for formation of A-T and G-C pairs. These Watson-Crick base pairings are illustrated in Figure 1.1.2 in which A forms two hydrogen bonds to T and G forms three hydrogen bonds to C.

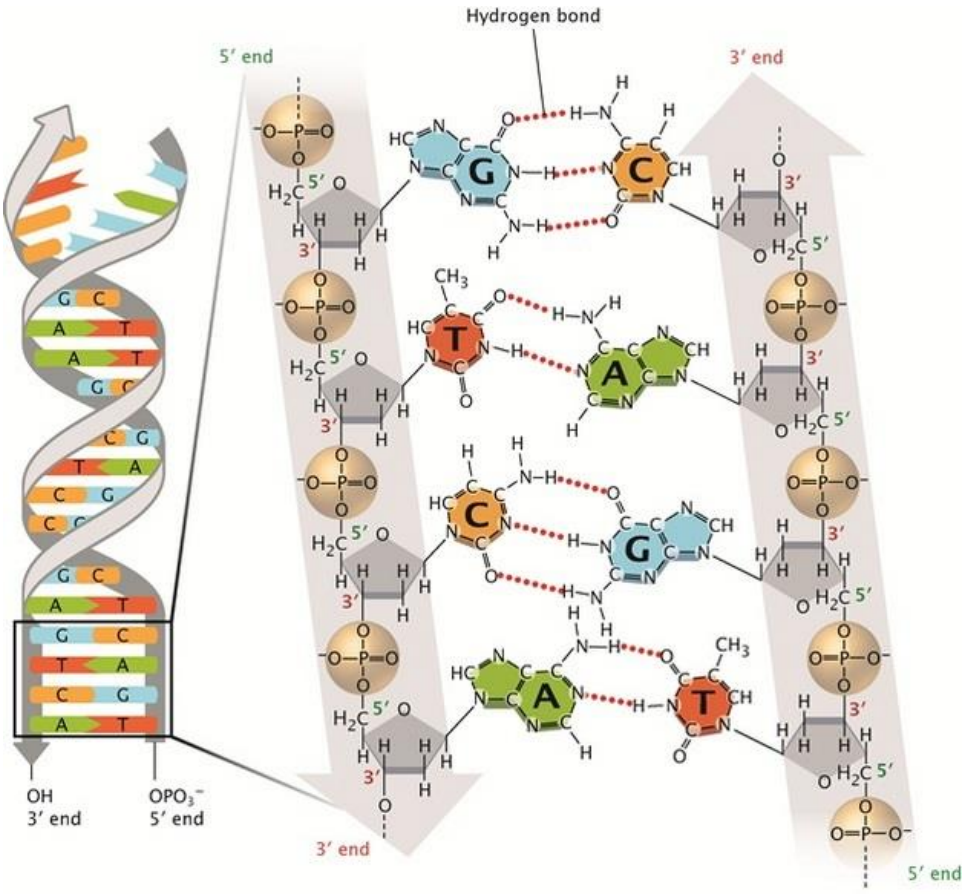


Figure 1.1.2. Illustration of Waston-Crick base pairing in a DNA helix. Taken from [1]

Watson-Crick base pairing are the predominant mode of base pairing in natural DNA; it is however, occasional mismatches as well as non-Watson-Crick base pairing such as either Hoogsteen or wobble base pairing[3-5] can occur. Other base pairing schemes and interaction have been reported for more complex DNA structures, such as triplexes[6, 7] and G-quadruplexes,[7, 8] but these only normally occur under very specific or narrow set of conditions. Specific base pairing or hybridization in oligonucleotides result in the secondary structure in oligonucleotides. While the double-stranded DNA (dsDNA) helix is the most familiar secondary structure of DNA, it is also possible for base pairings to occur within an individual sequence or in single-stranded DNA (ssDNA) due to self-complementarity within the sequences. Thus, unlike the common helix which is independent of the primary structures of complementary DNA, the self-complementarity of ssDNA sequences can lead to many unique secondary structures. While these secondary structures are less studied in ssDNA systems, various secondary structure are commonly observed in its ribose counterpart known as ribonucleic acid (RNA). While base pairing plays a crucial role in secondary structure formation, additional factors include van der Waals and hydrophobic interactions between bases, as well as π-π stacking effects. Repulsive electrostatic interactions due to the negatively charged phosphates can inhibit or weaken hybridization if not well shielded by ions from salt additions. Other factors that can weaken the stability of the secondary structure of oligonucleotides include extreme pH conditions (purine bases can be hydrolyzed at low pH[9] and the phosphate backbone can be degraded under high pH conditions[10]) and temperature[11].

Differences in the nature of the single secondary structure of an oligonucleotide can, in turn, effect the overall conformation or higher level tertiary structure of the

oligonucleotide. Depending on the oligonucleotides involved, the sequence content of pairing bases, and solution conditions, distinctive forms of the double helical structure can occur as shown in Figure 1.1.3 where the most commonly observed form of the double helix is the B-form for Watson-Crick base pairing of DNA. [7]

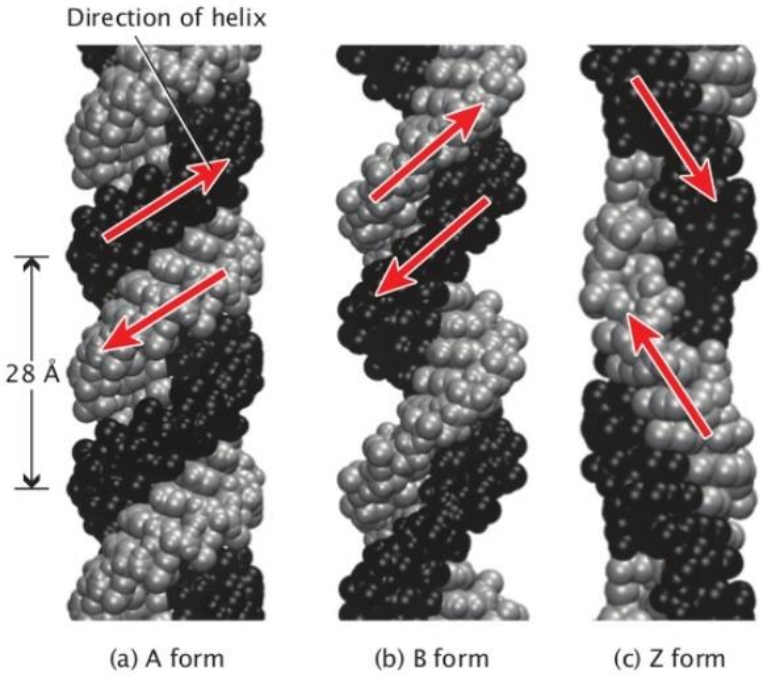


Figure 1.1.3. Helical conformations of dsDNA in which (a) is right-handed A form (b) is right-handed B form and (c) is left-handed Z form. Taken from [1]

The alternate A and Z forms have been shown to have implications in the biological functions of oligonucleotides as well as affect the overall stability of the formed duplex.[12]

1.1.1.2 Thermodynamic Properties of DNA and other Oligonucleotides

One metric to describe the stability of a given secondary structure of oligonucleotides is the value of the melting temperature (T_m), which is defined as the temperature at which half the total number of sequences are associated into their desired secondary structure. In other words, T_m is a measure of thermal stability of the secondary structure against dissociation. Melting temperature depends on the number of base pair matches, the composition of the base pair matches (ratio of stronger G-C to weaker A-T pairs), oligonucleotide sequence concentration, and ionic strength of solution.[13, 14] For example, since G-C pairs possess more hydrogen bonds than A-T pairs, sequences with higher G-C ratios result in higher melting temperatures. As a second example, increasing the total number of base pairs also tends to raise the melting temperature. The overall propensity for hybridization of sequences is described thermodynamically in terms of Gibb's free energy (ΔG), melting temperature (T_m), and dissociation constant (K_d) in which we consider the hybridization reaction between two complementary individual sequences to form double-stranded (ds) product as follows:

$$ssDNA_1(A) + ssDNA_2(B) \overset{K_a}{\leftrightarrow} dsDNA(AB) \tag{1.1.1.1}$$

where K_a is the equilibrium association constant. We can also describe this with the Gibbs free energy (ΔG) for this reaction as both

$$\Delta G^\circ = -RT \ln K_a \tag{1.1.1.2}$$

and

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ \tag{1.1.1.3}$$

Where R is the gas constant, T is absolute temperature, ΔH is enthalpy of the reaction, and ΔS is entropy. For oligonucleotides at the exact melting temperature value for a given oligonucleotide sequence, K_a=0.5 and T_m is defined as the temperature at which $\Delta G^\circ = 0$. After equating equations 1.1.1.2 and 1.1.1.3, one can rearrange terms to generate the following equation:

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ - R \ln K_a} \tag{1.1.1.4}$$

Finally, we can define the equilibrium dissociation constant K_d as the inverse of the equilibrium constant K_a:

$$K_a = \frac{1}{K_d} \tag{1.1.1.5}$$

By substituting K_a in 1.1.1.2 with equation 1.1.1.5 yields:

$$K_d = e^{\frac{\Delta G^\circ}{RT}} = e^{\frac{1}{R}\left(\frac{\Delta H^\circ}{T} - \Delta S^\circ\right)} \tag{1.1.1.6}$$

Since these equations are expressed in a generalized format, they can also be used to explain thermodynamics of self-interaction or self-hybridization of single-stranded (ss) oligonucleotides with themselves, e.g.

$$ssDNA(A) \overset{K_a}{\leftrightarrow} Self - Hybridization(A_{self}) \tag{1.1.1.7}$$

or even with a non-nucleotide species such as protein:

$$DNA(A) + Species(B) \overset{K_a}{\leftrightarrow} Complex(AB) \quad (1.1.1.8)$$

Since DNA has been widely studied by groups such as SantaLucia and Zuker[15-22], the predicted and experimental thermodynamic parameters of many inter-oligonucleotide and intra-oligonucleotide interactions are well documented, including values for nearest neighbour base stacking effects on ΔG°. This information enables computational methods such as Zuker's UNAFOLD web server to be employed to calculate predicted values of ΔG° and T_m for a given sequence or sequences under various in solution conditions through summation of experimentally determined values of ΔG for the 10 possible base pair nearest neighbour interactions:

$$\Delta G^{\circ}(total) = \sum n_i \Delta G_i\,(i) + \Delta G^{\circ}_{init\ G-C} + \Delta G^{\circ}_{init\ A-T} + \Delta G^{\circ}_{sym} \quad (1.1.1.9)$$

in which n_i is the number of incidences of a particular nearest neighbour pair, $\Delta G_i(i)$ is the free energy of the corresponding nearest neighbour pair (i), $\Delta G^{\circ}_{init\ G-C}$ and $\Delta G^{\circ}_{init\ A-T}$ are the free energy of a terminal G-C or A-T pair respectively, and ΔG°_{sym} is the free energy of self-complementarity.

Other considerations for the thermodynamics of oligonucleotides is the variations in enthalpy with temperature, since the assumption that enthalpy remains constant with respect to temperature (i.e. heat capacity, $\Delta C_p^{\circ} = 0$), is not always true. This temperature dependence can be expressed by the rearrangement of equations 1.1.1.2 and 1.1.1.3 to generate the van't Hoff equation:

$$\frac{\partial \ln K_a}{\partial(\frac{1}{T})} = -\frac{\Delta H^\circ}{R} \qquad (1.1.1.10)$$

or

$$\frac{\partial \ln K_a}{\partial T} = -\frac{\Delta H^\circ}{RT^2} \qquad (1.1.1.11)$$

From equation 1.1.1.10 and 1.1.1.11, it can be observed that K_a will vary with temperature unless $\Delta H^\circ = 0$. Determination of the dependence of ΔH° on temperature can be carried out by measuring K_a at various temperatures, taking the slope of K_a as a function of $\frac{1}{T}$ to yield $\frac{-\Delta H^\circ}{R}$. The second derivative of K_a as a function of T will yield ΔC_p. The value for ΔC_p can be estimated theoretically using methods in Zuker's UNAFOLD with $-T\frac{\partial^2 \Delta G}{\partial T^2} = \Delta C_p$.

In contrast to pure oligonucleotide systems, thermodynamic estimations for parameters in equation 1.1.1.8 involving the interaction of DNA or another oligonucleotide with a non-nucleotide species are not broadly characterized[23-26], as the number of variables such as the particular oligonucleotide sequence, non-nucleotide species, solution conditions, and modes of association are seemingly countless. While it is not possible to explore all potential DNA sequences interactions with a species, recent advancements in DNA solid-phase synthesis do allow for facile tailoring of both sequence and length, so investigation of many of sequences of interest are more possible.[27]

1.1.2 Other Candidate Oligonucleotides

1.1.2.1 RNA

Like DNA, ribonucleic acid (RNA) is also a biomacromolecule but has several key differences from DNA. Firstly, the backbone of RNA has a ribose sugar with has a hydroxyl group on the 2'-carbon rather than the 2'-hydrogen present in deoxyribose shown in Figure 1.1.2. In general, the presence of this 2'-hydroxyl on ribose results in lower chemical stability as compared to DNA, especially in alkaline and elevated temperature conditions (37°C<T<100°C).[28, 29] The 2'-hydroxyl can even result in self cleavage of the phosphodiester bonds in the backbone of RNA.[30] Another key difference from DNA is the substitution of the thymine base in RNA with a uracil base which is a chemical analogue of thymine with a demethylated 5'-carbon. RNA often occurs as a single-stranded biomacromolecule in nature, which results in greater observed structural diversity in contrast to dsDNA. In fact, ssRNA even possesses more conformational freedom than a comparative ssDNA sequence, which potentially allows for additional conformations (e.g. pseudoknot) not possible for ssDNA.

1.1.2.2 LNA

Locked Nucleic Acids (LNAs) are synthetic oligonucleotides that mimic aspects of both DNA and RNA. Similar to RNA, it possesses a ribose-like sugar with key modifications- the 2' hydroxyl group is replaced with 2'-oxygen and there is a methylene linker connects the 2'-oxygen to the 4'-carbon in the ribose as shown in Figure 1.1.4 to effectively lock the ribose into a C3'-*endo* conformation. LNA typically uses found in DNA (A, T, G, and C).

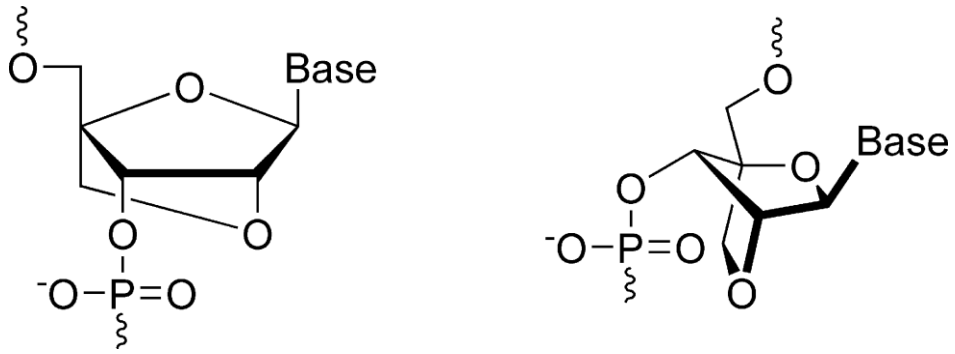


Figure 1.1.4. Chemical structure of LNA with the methylene linker (left) and in its C3'-*endo* conformation (right). Taken from [31]

These modifications of the ribose group results in more A-form helixes which are thought to strengthen base stacking interactions as compared to B-form dsDNA helixes. [32, 33] In addition to promoting stronger base stacking interactions, the methylene bridge also confers nuclease resistance, as compared to its RNA and DNA counterparts. [34, 35] The Milam group has demonstrated greater hybridization activity and fidelity of recognition between of complementary oligonucleotide sequences with LNA-based sequences compared to pure DNA sequences, though generally only modest differences were found in the association rate constants. [36] Other groups have reported similar findings in terms of overall affinity and fidelity of recognition of LNA to complementary oligonucleotides.

1.2 Aptamers

Aptamers are single-stranded oligonucleotides with high affinity and specificity for a particular non-nucleotide target. Using *in-vitro* screening processes, aptamers have been identified for many targets ranging from ions to small molecules to proteins and even whole cells. [37-44] Aptamers are commonly thought of as the oligonucleotide analogues of antibodies, but with many significant advantages over their antibody counterparts. Similar

to antibodies, aptamers have been characterized by their equilibrium dissociation constants, and have exhibited values of K_d typically in the picomolar to the micromolar range. [45] While antibodies typically rely on animal hosts for their *in vivo* generation, aptamers are generated from a synthesized oligonucleotide library which enables easy customization of base content and oligonucleotide modifications with lower cost and high reproducibility during their *in vitro* selection processes. [45-48] It is largely accepted that aptamer-target binding is strong, but non-covalent and occurs through structure complementarity between self-hybridized aptamer and its target; this structural complementarity arises from several factors including the aromatic rings involved in base stacking, electrostatic and van der Waals interactions, hydrogen bonding or any combination thereof.[49] Aptamer interactions with their target are often treated as a "lock-and-key" binding interaction similar to that of an antibody and its antigen, but other studies indicate that other mechanisms are involved in binding[50, 51], discussed in detail in section 1.5. These properties of aptamers have inspired large-scale investigation for uses in therapeutics, biosensing, diagnostics, and research.

1.3 Conventional Aptamer Screening

As stated previously, three labs independently and simultaneously developed a similar procedure for *in-vitro* selection of aptamers in 1990. [52-54] Robertson and Joyce used their *in-vitro* selection procedure to select a variant of *Tetrahymena* ribozyme to cleave single-stranded DNA more efficiently than the naturally occurring wild type of the enzyme. [52] Through repeated cycles of mutagenesis and selective amplification, the first RNA sequence/aptamer to be generated *in vitro* to specifically cleave single-stranded DNA was identified by the Joyce group. At the same time, Tuerk and Gold used a randomized eight-

base length region within an RNA sequence that was previously known to bind to the bacteriophage T4 DNA polymerase. Through progressive rounds of selection and amplification of the sequence, Tuerk and Gold demonstrated that they could identify a new overall sequence with improved binding capabilities to the T4 DNA polymerase.[54] It is from this publication by Tuerk and Gold that the eventual preferred descriptive name for this *in vitro* selection process originated: "Systematic Evolution of Ligands by Exponential Enrichment" (SELEX). The third group to publish a related process the same year was Szostak's group. Unlike the previous groups, Ellington and Szostak utilized a completely randomized RNA library which did not contain any previously known RNA sequence or species with affinity for their chosen target, a molecular dye.[53] This non-preferential selection is what most current SELEX processes use. [45] Ellington and Szostak were also the first to coin the term "aptamer" in this paper, using the Latin "*aptus*" meaning "to fit" and "-*mer*" denoting a polymer-like repeat unit to describe their identified RNA sequence which preferentially bound to their molecular dye target, in what they hypothesized to occur through "fold[ing] in such a way as to create a specific binding site for small ligands." [53] With these landmark papers as a foundation, the SELEX-based selection processes has become wide-spread in its use to identify aptamers sequences or oligonucleotide based ligands comprised of single-stranded oligonucleotide sequences with high affinity and specificity for their particular non-nucleotide target.

1.3.1 Conventional in vitro *SELEX*

The first step in preparation for SELEX is the generation of a random library of oligonucleotides such as DNA that will serve as a "pool" of potential aptamer sequences during panning. A typical DNA library consists of a large number of ssDNA fragments

(~10^{15} molecules) that share a templated design comprised of a central random region of 20-80 nucleotides with flanked 5' and 3' ends that are specific or fixed base segments.[45] These fixed end segments facilitate amplification of the sequences throughout the SELEX process as short complementary sequences called primers are designed to hybridize to these fixed end segments and allow PCR to proceed as shown in Figure 1.3.1.

`

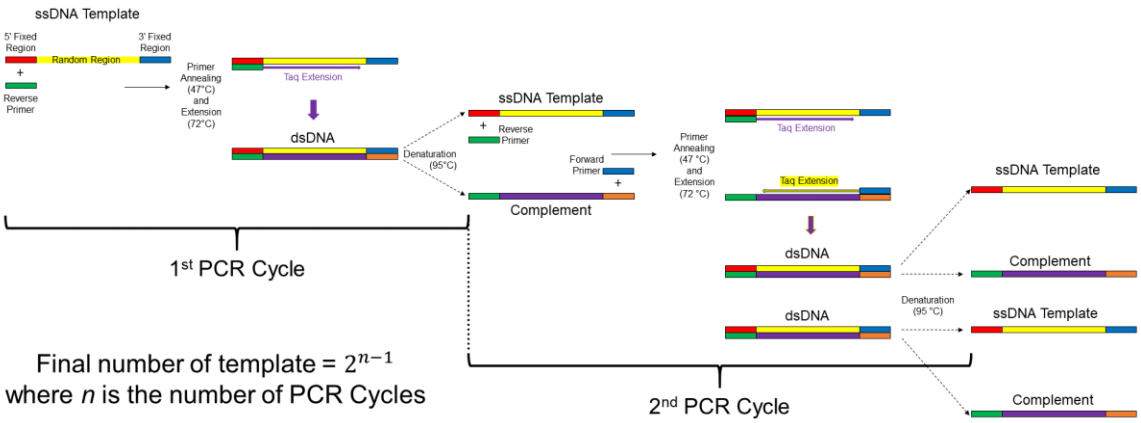


Figure 1.3.1. Illustration of the ideal polymerase chain reaction for a single ssDNA template with a random region. A PCR cycle is comprised of (1) annealing of the primers at 47°C (2) extension of the primer by *Taq* polymerase at 72°C in the presence of monomeric species called dNTPs (not shown) and (3) denaturation of the double-stranded DNA at 95°C.

While there are ~10^{15} ssDNA in a library, there are typically multiple copies of each sequence, so the actual number of unique sequences may range from ~10^{8}-10^{12} sequences. If the ssDNA library is later to be transformed into an RNA library, a sense primer on the 5' end containing the T7 promoter sequence is necessary along with the antisense primer. [45] The T7 promoter sequence allows the T7 RNA polymerase to transcribe the ssDNA to RNA, however if a ssDNA library is to be used no additional preparation is necessary. If performing SELEX with an RNA library, each round requires the RNA to be reverse-

transcribed back into ssDNA, subsequently amplified via reverse transcription polymerase chain reaction (RT-PCR), and transcribed back into RNA with T7 RNA polymerase. [45] DNA continues to have several advantages over RNA for its use as an aptamer, even ignoring these extra steps during PCR that are required for RNA SELEX; DNA possesses greater chemical stability in *in vivo* environments as compared to RNA and thus makes it a more attractive ligand option.

After preparation of the oligonucleotide library, a common precursor step to SELEX is counter selection. Counter selection is essentially a partitioning step involving a negative target that reduces the initial library size by removing sequences that have affinity to undesirable targets such as reaction vessels, immobilization substrates, side products of a reaction with the desired target, a closely comparable species to the desired target, or even a different form of the desired target. In Figure 1.3.2, a counter selection step is illustrated for the case in which the desired target is immobilized onto a substrate, so a counter selection against the substrate is a desirable step since aptamers with a significant affinity for the substrate are not a desirable aptamer candidate. It is possible to employ multiple counter selection steps to enable more specific selection of aptamer candidates for the desired target. [42, 45, 55]

SELEX itself begins with an incubation step to introduce the library to the target under a desired set of environmental conditions for promoting specific binding. Parameters include choice of buffer salt and pH conditions, temperature conditions, substrate for convenient immobilization of small targets, and addition of potential blocking agents to promote for equilibration of the library with the target. Following incubation, the next step in the SELEX round is partitioning, in which unbound sequences are separated from target-bound

sequences. Depending on parameter choices, such as a substrate for target immobilization, this step can be potentially accomplished via magnetic bead separation, centrifugal separation of a bead substrate, gel shift assay separation, a wash of the target functionalized solid phase in HPLC, or nitrocellulose filtration. [55] After partitioning, the remaining bound sequences are eluted by changing buffer conditions and/or temperature to promote dissociation of target-sequence complexes, and the eluted DNA sequences are subsequently amplified via PCR as shown in Figure 1.3.1 or via RT-PCR for eluted RNA sequences. The previously bound and now amplified copy numbers of these sequences are

used to enrich the pool of candidate sequences and can now be reintroduced to the desired target in a new SELEX round as shown in Figure 1.3.2.

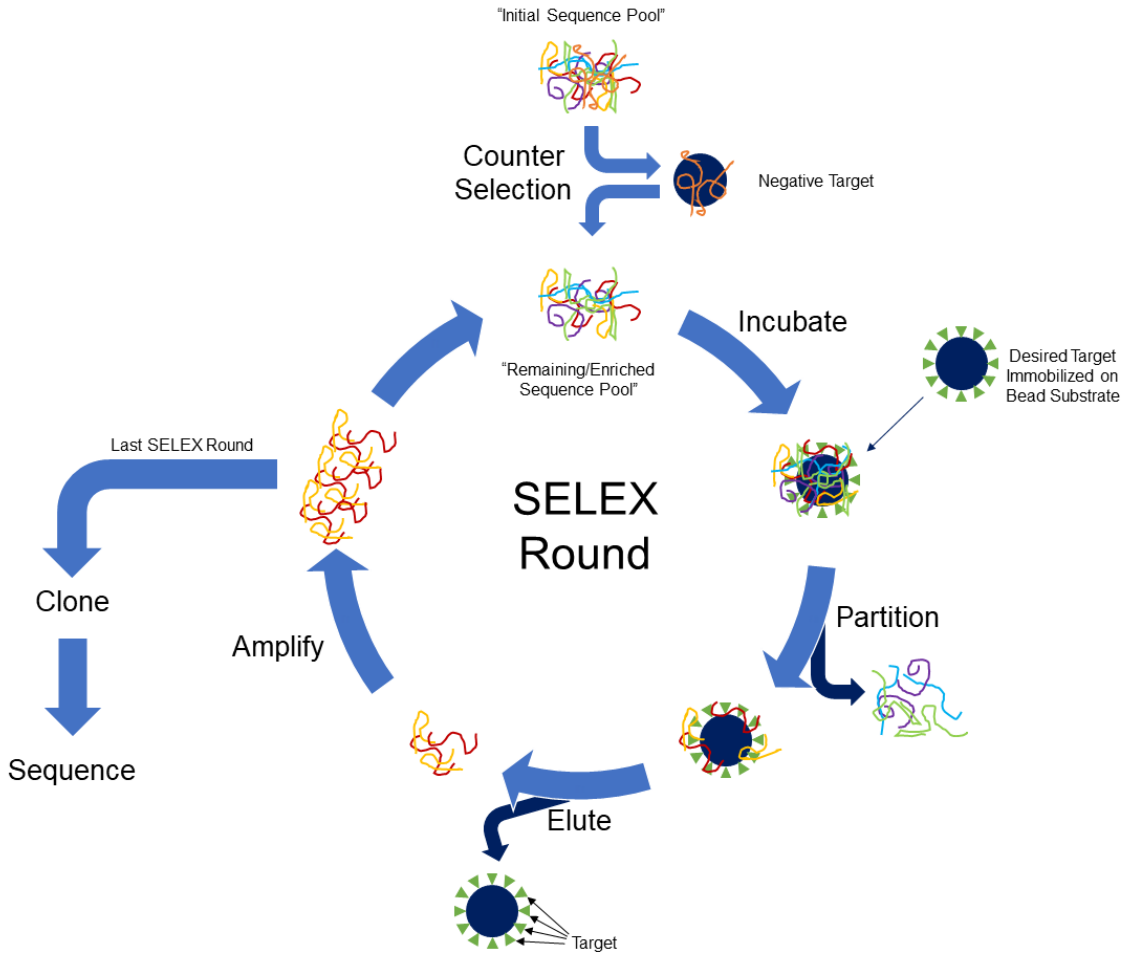


Figure 1.3.2 Schematic of conventional SELEX. One round of SELEX includes (1) incubation of the random oligonucleotide pool with the desired target (2) partitioning of non-binding sequences (3) the elution of bound sequences from the target (4) PCR based amplification of the previously bound sequences to be used in the next SELEX cycle. The counter selection illustrated is an optional precursor step to remove sequences that bind to

a non-desirable target before they are introduced to the desired target. After a set number of rounds of SELEX, cloning and sequencing take place after a final PCR step.

Rounds of SELEX are typically repeated multiple times in order to ideally promote the selection of a set of sequences with the highest affinity out of the library pool. Depending on the number of SELEX rounds performed (normally between 10-30 rounds), complexity of the target, and the stringency of selection conditions, the SELEX process can take approximately two months to complete.[42, 45] Some more recent studies suggest that while there are relatively few sequences with high affinity in a given pool of ~10^{15} sequences, [56] identification of these high affinity sequences can be accomplished in as few as three SELEX rounds by employing high throughput next generation sequencing (NGS) rather than conventional, low throughput sanger sequencing.[57] Following the final SELEX round, the final remaining sequences are typically cloned a subset of clones and sequenced. As mentioned previously, more modern methods skip the cloning step, and instead sequence the entire remaining pool with NGS rather than just the smaller subset available with sanger sequencing.[58]

Post-SELEX analysis of the identified sequences commonly includes primary structure analysis, or predicted secondary structure analysis, and possibly, quantitative binding affinity studies. The goal of primary structure analysis is to pinpoint any shared or common base segment(s) requisite for binding to the target, while the secondary structure analysis attempts to identify any common secondary structures of these sequences or self-hybridized segments that may bind to the target. Since primary structure directly correlates with secondary structure, both shared base segments and common secondary structures of resulting sequences are referred to as consensus motifs, and often treated as potential

binding motifs. [58, 59] For consensus motifs within base segments, it is common to ignore the fixed end portions in the sequence template as they occur in every sequence, but work by *Ouellet et al.* gives credit to the postulation that fixed regions interference with formed structures should still be considered for secondary structure motifs. [60] Binding affinity studies to determine equilibrium dissociation constants (K_d) are common but can require specialized equipment, labelling schemes, and differing conditions and experimental setup from SELEX. Section 1.5 will address K_d measurements in more detail.

1.3.2 Pitfalls and Limitations of Conventional SELEX

While SELEX discovery of aptamers has a number of advantages over *in vivo* methods to generate antibodies for a target, conventional SELEX does have a number of limitations and pitfalls. One limitation of conventional SELEX is that not all molecules or substrates are suitable targets. Molecules that cannot be produced with high purity or sufficient quantity should not be considered, as enrichment of non-functional or nonspecific sequences is more likely. Even for targets with high purity and yield, there is potential for enrichment on nonspecifically binding sequences through a number of pathways which include the following: positively charged species which can nonspecifically bind to the negatively charged backbone of oligonucleotides, species with hydrogen bonding candidate sources which can nonspecifically bind to nucleic acid bases, and hydrophobic domains such as aromatic compounds which can also nonspecifically bind to bases.[61, 62] Conversely, highly negatively charged species can be challenging as targets due to electrostatic repulsion with oligonucleotides. As desired targets shift to more complex systems, such as living, diseased, cells, the increase in potential, but undesired, binding sites (e.g. a receptor found on both normal and diseased cells) can greatly affect the degree

of specificity achieved in resulting aptamers.[63] To ensure these issues are not generating false positive hits among candidate sequences, counter selection becomes key. Additionally, some authors have even postulated that the similarity in structures observed for distinct aptamer sequences for the same target (in separate SELEX processes run by different groups) “suggests that some factors can affect the direction of aptamer evolution which may result in the failure of aptamer selection for specific target or site of interest.”[47] This directed aptamer evolution may in part be due to interference or promotion of binding events played by fixed base segments.

A second set of limitations on SELEX involves the challenge to uniformly standardize SELEX across the community. Since targets vary, an aptamer selection protocol appropriate against one target may not be possible for other targets due to the incompatibility in immobilization and separation schemes; target size, availability and stability under selection conditions; and the lack of a suitable counter selection or negative target. An additional factor relating to this limitation is the environmental stability of the oligonucleotide in a complex environment such as serum due to the presence of oligonucleotide-cleaving nucleases, destabilizing pH conditions, presence of other nucleic acids, and hydrophobic and/or positively charged species (e.g. lysine residues in a protein) in a biological environment. Thus, oligonucleotides may be degraded or have structure loss during the selection process itself. This set of limitations results in SELEX requiring a significant amount of optimization for most biological targets. Though automated SELEX methods exist[57, 64, 65], the challenge to incorporate specialized robotic systems means that manual handling (and all its associated errors) is currently used and likely to continue for the immediate future.

Lastly, a very specific pitfall of SELEX is the quality of the synthesized library and subsequently selected pools in each round of SELEX. In 2005, Musheev and Krylov published their findings detailing differences in PCR of a ssDNA template with a random central region versus a homogenous sequence population which demonstrated that PCR for random libraries proceeds markedly differently and requires significant optimization to reduce undesirable side product formation.[66] Since then multiple groups have continued to report bias in PCR during SELEX in both base content and side product formation using either non-equilibrium capillary electrophoresis (NECEEM) or NGS characterization and have proposed models of by-product formation and resulting selection bias. [67-69] Recently, another group reported that bias arises even in the ordered template from different manufactures, persists in PCR, and can affect the overall affinity of the identified aptamers. [70] Emulsion PCR (ePCR) [69, 71] as well as now droplet digital PCR (ddPCR) has now been shown to be one method to minimize PCR bias and side product formation in random oligonucleotide libraries. [70]

1.4 Aptamer Targets

This next section provides an overview of some of the general trends in target choices for SELEX-based aptamer screenings. As stated previously, aptamers have been selected for targets ranging from ions and small molecules to proteins and even whole cells. [37-44] The next question can entail the actual application of an aptamer. The answer is that an aptamer may potentially be used as both a recognition, detection tool as well a potential therapeutic tool. [50, 72-74] As examples, a number of potentially therapeutic and diagnostic aptamer sequences have been investigated as capture agents for: human immunodeficiency virus (HIV) related proteins, Hepatitis C related proteins, leukemia

cells, immunoglobulin E, thrombin, and vascular endothelial growth factor (VEGF).[45] The first FDA approved aptamer therapeutic is known as Pegatnib or Macugen, a pegylated aptamer for VEGF.[75]

1.4.1 A Model Aptamer Target: VEGF

VEGF is a basic heparin binding glycoprotein which is encoded by a single gene and can be expressed in four different isoforms: $VEGF_{121}$, $VEGF_{165}$, $VEGF_{189}$, and $VEGF_{206}$, in which the VEGF isoform number is equal to the number of amino acid residues present. [76, 77] The most common isoform expressed in humans is $VEGF_{165}$ (~45 kDa). While the $VEGF_{121}$ isoform is acidic, the other isoforms are basic, and exhibit higher binding affinity to heparin, a glycosaminoglycan which functions as an anticoagulant. [78] $VEGF_{121}$ is a freely soluble protein, and though $VEGF_{165}$ is somewhat soluble, a significant portion of the isoform remains physically associated with the cell surface or extracellular matrix when secreted.[79] VEGF assumes an antiparallel heterodimer structure and possesses two binding domains, the receptor-binding domain and the heparin-binding domain.[80, 81] Correlation of the structure of $VEGF_{165}$ with its function has been effectively demonstrated through limited proteolysis of VEGF which results in a 7-10 fold reduction in its bioactivity. [82] The high affinity binding of VEGF (K_d ~1-10 pM) to multiple receptors also has been widely characterized.

The unique structure and properties of VEGF extend to its distinctive roles as a mitogen and angiogen. VEGF has been described as a "potent" mitogen with an effective dosage that produces a therapeutic response or ED_{50} as low as 2 pM. [77] As implied in its name, groups have reported VEGF selectively induces mitosis or mitogenesis of endothelial cells

from veins, arteries, and lymphatics, but not for other cell types. [83-89] As an angiogen, VEGF regulates angiogenesis through interactions with tyrosine receptor kinases which can result in either desired physiological angiogenesis (e.g. tissue regeneration) or undesired pathological angiogenesis (e.g. tumor growth). [90, 91] In physiological angiogenesis, VEGF has been shown to play crucial roles in embryonic vasculogenesis and angiogenesis,[92, 93] postnatal development,[94, 95] skeletal growth and bone formation,[96, 97] and ovarian angiogenesis.[97, 98] Conversely, pathological angiogenesis has been found to be mediated by VEGF in solid tumors and hematologic malignancies,[99, 100] intraocular neovascular syndromes,[101-103] inflammation and brain edema,[104, 105] and pathology of the female reproductive tract. [106, 107] Thus, the dual VEGF plays in mediation of angiogenesis in a variety physiological context makes it an ideal target for therapeutic applications as both a target for inhibition[108] or promotion of angiogenesis activity. [109, 110]

The trifecta of its unique molecular structure, structure related function and relevant therapeutic and diagnostic applications allowed VEGF come into the spotlight as an ideal target for aptamer identification through SELEX since an aptamer potentially enables tailored control of inhibition or uptake of a molecular species as well as potentially function as a delivery platform. [74, 111-113] The first reported and subsequent aptamers for $VEGF_{165}$ were identified by Janjic and coworkers. [114-116] Janjic *et al.* continued to investigate affinity and structural modification for stability and truncation of the RNA aptamer for $VEGF_{165}$ and later demonstrated that they could in fact inhibit receptor binding of VEGF as well as subsequent VEGF mediated vascular permeability through targeted aptamer binding to the receptor for VEGF. [117] After clinical trials, a pegylated version

of the RNA aptamer for VEGF165 was approved by the FDA for use in humans under the name Pegatnib or Macugen.[75, 90] Since the identification of RNA aptamers for $VEGF_{165}$ by Janjic *et al.*, other groups have identified additional DNA aptamers for $VEGF_{165}$.[118] The aptamers for $VEGF_{165}$ interactions have continued to be investigated by additional groups to study aptamer-VEGF binding mechanisms, affinity and kinetics after selective modification, truncation, and deletion of portions of the aptamer structure. [119, 120] The Milam group has explored regulated uptake, release, and regeneration of aptamer binding capability using competitive target displacement strategies. [81] Thus, VEGF has served as a model target for selection of aptamers via SELEX to help pave the way for aptamer-based therapeutic applications.

1.4.2 Modification of Aptamers for Optimization of Target Selection

Several of the previously discussed limitations of natural nucleic acids (i.e. DNA and RNA) can be mitigated by chemically modifying oligonucleotides to alter their physical, chemical, or structural properties. The following sections will discuss three overarching schemes for optimization of aptamer performance through the following oligonucleotide modification routes: template design, synthetic nucleic acid incorporation into library pool, and post-SELEX sequence modification.

1.4.2.1 Template Design Modification

One of the most easily implementable modifications to oligonucleotide libraries is to bias their properties even before introduction to the target through thoughtful design of the template strands. There are several possible stratagems to even adapt the template design to better suit a specific target, such as (1) biasing base content in the random region; (2)

tailoring the base length of the random region; (3) redesigning fixed base segments; and (4) intentionally introducing conserved secondary structure into the library.

Biasing base content in the random region can allow for formation of more unique secondary structures, enable better chemical stability in the screening environment, or promote selection of higher affinity from candidates possessing bases with a known affinity to a specific target. This has been effectively used with G/T and G/C rich libraries to promote G-quadruplex formation. [121] This G-quadruplex featured library is generally associated with higher serum stability and cell uptake.[121] As described in Chapter 3, the Milam group intentionally utilized an A-rich library for panningagainst gold-based target based on prior reports of the stronger interactions between gold and adenine. [122] Other groups have subsequently biased their chosen sequence after selection in which, for example, an aptamer rich in A/T was originally identified for an A/T binding protein, and then mutated into a non-A/T rich sequence to demonstrate the resulting reduction in affinity. [123] Though the motivation for base bias has merit, one must exercise particular caution since biased weighting of bases in the random region also may be an undesired side product from unoptimized PCR bias.

The second option involves tailoring the base length of the random region. To this purpose, most have argued for significant reduction of the random segment of the template as longer random region templates cannot potentially display all their variations within a given library of 10^{16} or fewer total sequences.[124-127] On the other hand, while shorter random regions have been successful for identification of aptamers [124, 125, 127, 128], these shorter base length libraries may not be able to achieve sufficient structural diversity in terms of secondary structure to yield high affinity for the target. [126] To balance these

competing factors, Wang *et al.* have proposed a scheme in which the starting template has a short random region, but after every SELEX round the random region is extended.[129] Another very recently developed compromise to this issue has involved stepping libraries which use an array of random region lengths.[130]

Rather than modifying the entire template strands, modification of the fixed base segment has been a optimization strategy since the development of PCR.[131] Shortly after the introduction of PCR as a sequence amplification technique, the importance of primer design and its complement became quickly apparent.[132] Furthermore, these fixed base segments can affect aptamer selection due to their potential direct interaction with either the target itself or with the random region.[60, 133] The Krylov group investigated modifying the base length of the fixed base segments for the purposes of optimization of PCR of library templates for SELEX. [66] There is common agreement that shortening the fixed base region is ideal. [134] One group has even proposed eliminating fixed regions entirely through development of new methods for a cycle consisting of: removal of fixed regions from the library, incubation with the target, and subsequent regeneration of the fixed regions. [133] A simple solution for preventing fixed region interference for some targets has been to just introduce the complementary sequence to the fixed regions during incubation with the target to promote the formation of duplex ends, so the fixed regions will be ideally be associated in a duplex that do not compromise binding activity of the random region with the target.[60] Another recent modification to the fixed base segments also involves introducing a truncated complementary sequence to any one of the fixed regions, however in this instance the complementary sequence itself is immobilized to a substrate. [135]

The last template design optimization scheme is the intentional introduction of conserved secondary structure into the library template. To our knowledge, the first to propose this scheme was Szostak's group, in which a conserved central hairpin loop structure flanked by a random region on each side was used to facilitate identification of an aptamer for GTP. [136] Other groups have also reported successful identification of aptamers utilizing conserved structures to enhance hairpin loop formation within the central random region [133, 137] Another proposed concept is that of structure-switching aptamers by Nutui and Li who also used a conserved central structure flanked by random regions but utilized fluorescently labeling on complementary sequences to either of the fixed ends and another complementary sequence to the central structure with a quencher.[138] This integration of dye into the structure of the template design enabled identification of aptamers in which the target could displace the central complementary sequence to enable signaling of any binding activity. Secondary structure conservation in the template should be thoughtfully introduced since self-complementarity of the fixed regions may also occur to affect target-aptamer binding. [60, 136, 137]

1.4.2.2 Synthetic Nucleic Acid Modification

While DNA and RNA possess natural diversity as aptamer candidates, synthetic modification of nucleic acids embodies the opportunity to derive new aptamer species with potential for higher affinity and degree of specificity. Previously the main barrier to implementation of synthetic nucleic acid modifications was the lack of compatibility with DNA and RNA polymerases, but recent advancements have enabled discovery of compatible modifications with identification of compatible polymerases for these

techniques.[139] There are several key locations for possible chemical modification of nucleic acids: the backbone, the sugar, and bases.

There are currently three major possible modifications of the backbone: phosphorothioate, boranophosphate, and phosphonate, as shown in Figure 1.4.1(a). Successful SELEX identification of aptamers has been performed with phosphorothioate[140-143] and boranophosphate backbones.[144] A recent literature search for aptamers with phosphonate backbones did not yield any examples, but other phosphonate groups attached to aptamer 3' end has been explored as a method to inhibit neutrophil elastase.[145, 146] Benefits of using phosphorothioate backbones include enhanced nuclease resistance[141] and immunogenicity.[147] Additionally, phosphorothioate backbones can potentially allow for higher affinity binding through formation of additional unique structures not possible with DNA and RNA.[140, 143] For proteins exhibiting cysteine residues or other biomolecules with thiol or sulfate chemistry, disulfide bond formation is also possible with a phosphorothioate (e.g. heparin).[141] Boranophosphate backbones in aptamers have the potential advantage of hydrophobic, sterospecific, and polar properties compared to the phosphate backbones which could enhance affinity; however, SELEX for some targets with this chemistry has proven difficult.[144] Despite potential challenges, the Shaw group has previously reported successful identification of an boranophosphate aptamer to ATP.[148]

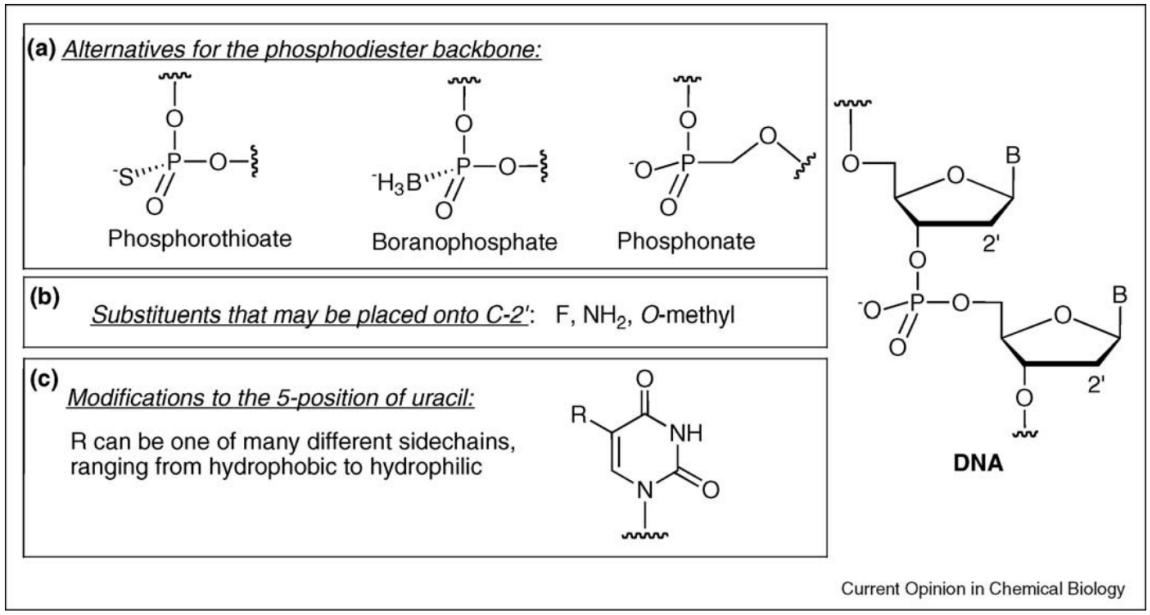


Figure 1.4.1 Chemical modifications to DNA tolerated by polymerases where (a) backbone modifications (b) sugar modifications and (c) uracil modifications are shown. Taken from Ref. [139]

Many sugar modifications are possible to create xeno-nucleic acids (XNAs) where X denotes the modification[149] as only partially shown in Figure 1.4.1(b). The XNA of greatest interest for the purposes of this review is LNA with the 2'-methylene and has previously been discussed as an oligonucleotide in section 1.1.2.2. While LNA does hold great promise in SELEX, LNA has most commonly been used as a modification post selection or integrated into only the fixed regions during selection.[149] Thus far, implementation of LNA in the random region during selection has only been achieved by Kasahara *et al.* who reported a chimera LNA aptamer for thrombin with a K_d of ~20 nM. [150]

The final candidate portion of nucleic acids for modification are the bases present. As shown in Figure 1.4.1**(**c), the uracil base contains a 5'-carbon that can be chemically modified while still capable of recognition by polymerases used for PCR.[139] The number

of potential modifications to bases and synthetic bases for SELEX are numerous; with 24 different base chemistry modifications reported, and of which 12 chemical base modifications are reported to be polymerase compatible.[151] To date, SELEX using these modified oligonucleotides has not been reported however, this large number of base modification possibilities along with accompanying variation in chemical and physical properties pose strong potential for customizing a library to a particular target.

1.4.2.3 Post-SELEX Primary Structure Modification

The last scheme for optimization of aptamers for a specific target involves post-SELEX primary structure modification. This modification scheme can be described as belonging to one of four categories: truncation (of base segment(s)), deletion (of individual bases), insertion/extension. These methods are commonly employed to attempt to reduce an aptamer sequence to only its essential binding structure, elicit binding motif information, and generate higher affinity for a target or greater chemical stability in harsher environments (e.g. *in vivo*). Characterization binding of VEGF aptamers have effectively used each of these approaches. Truncation of VEGF aptamers has been successfully used to identify shorter sequences that still possessed high affinity to $VEGF_{165}$.[114, 120] Individual base deletion in VEGF aptamers has enabled identification of minimal aptamer structure;[117] Extension of VEGF aptamers has been employed to stabilize secondary structure formation to promote higher target affinity.[119] Finally, base insertion of positional mutations has also enabled high affinity in VEGF aptamers. [152] Though success has been reported, the candidate parameter space (e.g. which base to alter, etc.) is often very large and interdependent. Moreover, because post-SELEX modification only

occurs after aptamers have been selected, it has only minimal utility in adapting the SELEX process to other target species.

1.5 Characterization of Aptamer Target Binding

A key question that remains to be explored in literature is the exact nature of binding of aptamers to their target. Since aptamers evolve from different libraries in various conditions for unique targets and SELEX protocols are not uniformly implemented across the community, there is no singular answer. To address this issue, several models have been developed along with techniques and methods to characterize an aptamer binding to its target in addition to evaluating the aptamer sequence and predicted secondary structure. The following sections will cover models for aptamer target binding, determination of the equilibrium dissociation constant, and motif discovery.

1.5.1 Models for Aptamer Target Binding

Since aptamers, whether natural or modified oligonucleotide, are a relatively new class of biomacromolecules, initial models for their interaction with other species have been based on previous models for other molecular interactions. The first and still most prevalent simplistic model used to describe aptamer-target complex formation is the “lock and key” model which derived from Emil Fischer’s early descriptions of enzyme-substrate interactions and has also been used to describe antibody-antigen binding. [153, 154] The lock and key model assumes that an aptamer is specific to molecules of a certain shape, as in the analogy of the molecule representing the lock and the aptamer representing the key that fits in the specific binding pocket of its matching lock. While the analogy is simple to envision, it may not completely describe the mechanism for aptamer target binding.

Shortly after coining the term aptamer, Szostak's group described the formed folded structure of ssDNA as playing an essential role in binding to its target. [155] This description, with later confirmation by others, formally led to the model used to describe most aptamer-target systems. As shown in Figure 1.5.1, the conformation state of the aptamer results in its recognition-based binding of its target.[45]

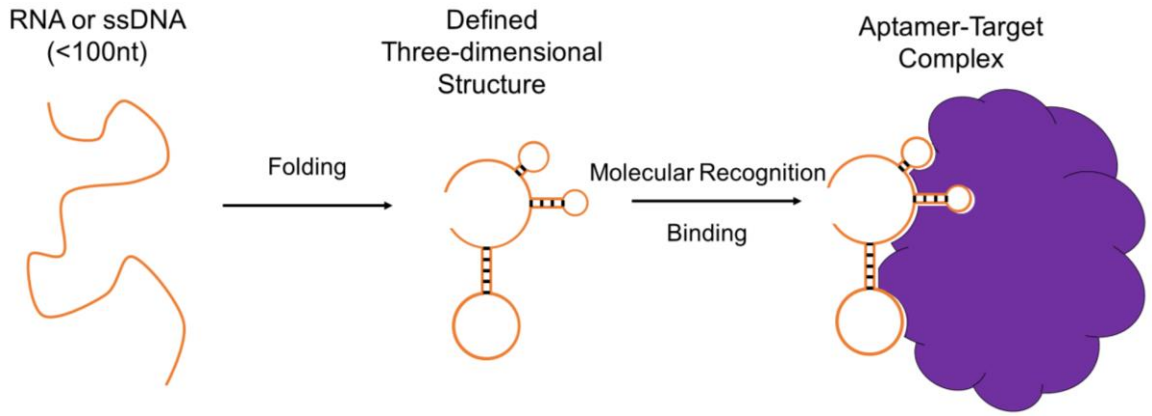


Figure 1.5.1. Illustration of the conformational selection for aptamer-target complex formation. Adapted from Ref. [45]

While this model is popular, some aptamer systems have been shown to deviate from this binding mechanism. [49, 156] Structure-switching aptamers in which conformational changes occur in aptamers upon target binding, for example, would not be possible under this model.[138] Thus, an alternative pathway for aptamer binding is similar to a theory that was first proposed by Koshland in 1958 for enzymes, namely an induced fit. [49-51, 154] Under this model, multiple potential pathways for aptamer target binding can be diagrammed similarly to the interactions between an enzyme and its substrate, as shown in Figure 1.5.2.[157-159]

From Figure 1.5.2, two possible pathways for aptamer binding can be observed, namely: one for conformational selection and another for induced fit. For conformational selection, an aptamer first folds into its self-complementary structure, forming a defined three-dimensional structure (though it is typically represented as a two-dimensional secondary structure) to enable recognition and binding to its target. In the induced fit pathway, the aptamer first encounters its target and then dynamically folds to bind with the target. While these binding pathways may seem exclusive, it may be possible for elements of both pathways to occur, as reported in non-covalent protein-protein interactions. [160] For simplicity, the shown binding model only assumes conformational selection and induced fit of the aptamer species, while the target's structure is assumed to be static.

To more deeply understand the nature of aptamer-target binding, it is important to acknowledge these various aptamer-target binding models to probe into the origins of affinity generation as well as the role of any motifs in binding. Despite the proposal of this newer induced-fit model, most publications still presume the "lock and key" model sufficiently describes aptamer binding equilibria and kinetics.

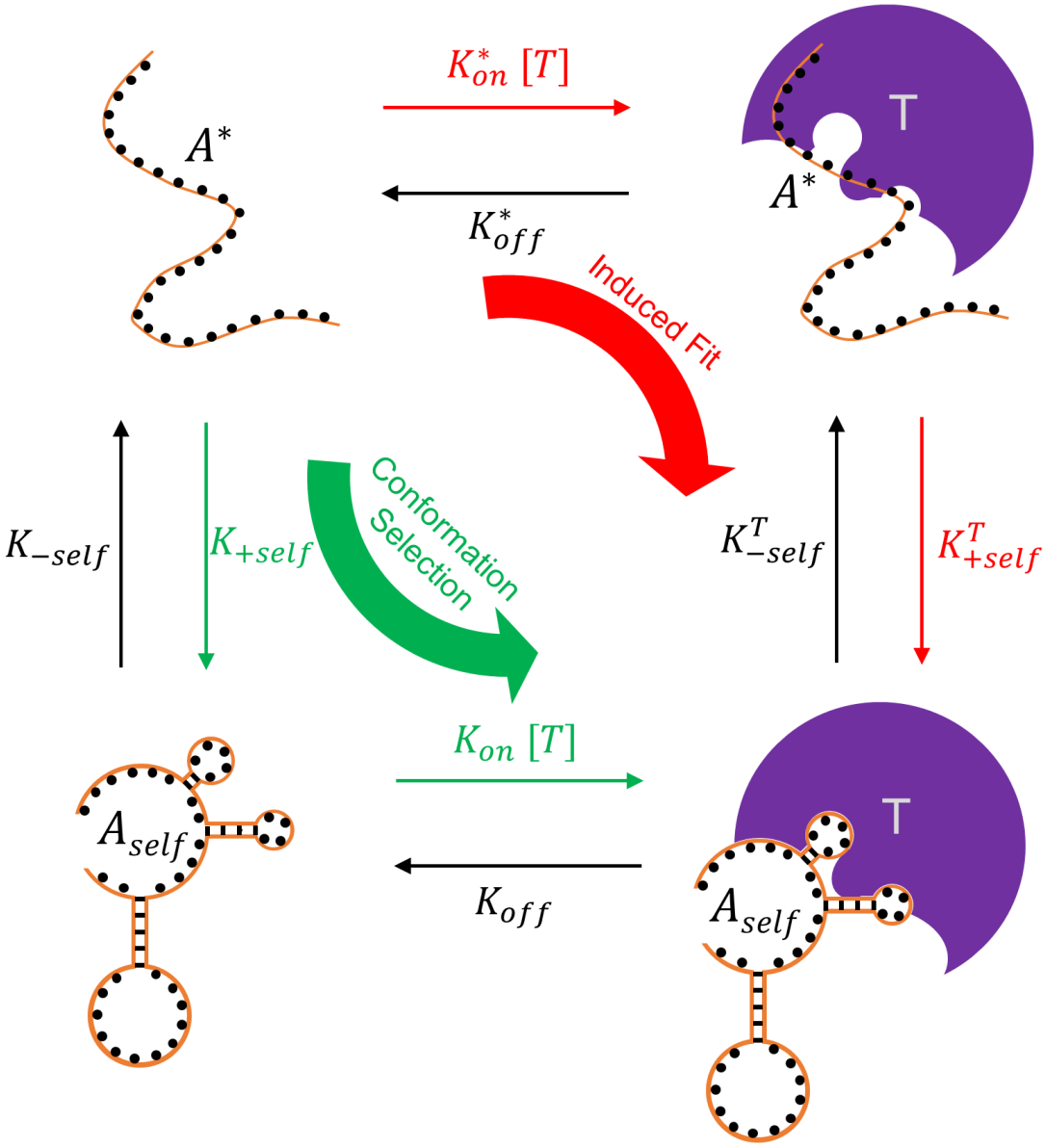


Figure 1.5.2. Reaction pathways for four species: (*A**) is the unfolded aptamer, (A_{self}) is the self-folded aptamer, (*A*T*) is the unfolded aptamer-target complex, and ($A_{self}T$) is the self-folded aptamer-target complex. Red arrows represent induced fit pathways while green arrows represent conformational selection pathways. Schematic adapted for aptamers from [158]

1.5.2 Determination of K_d

The most common quantitative metric of binding affinity in aptamer systems is the equilibrium dissociation constant, K_d. This section will overview one of the most common fitting models for K_d measurement as well as briefly summarizing the most common analytical methods used to measure K_d.

1.5.2.1 Michaelis-Menten Binding Curve Fitment for Aptamer K_d

While the Michaelis-Menten equation was originally developed for enzyme-substrate kinetics, it has been adapted to aptamer-target system presumed to follow a lock and key-type binding model description starting with equation 1.1.1.8 from section 1.1.1:

$$DNA(A) + Species(B) \overset{K_a}{\leftrightarrow} Complex(AB) \tag{1.1.1.8}$$

We will reduce this to just symbols where A represents the unbound aptamer, B represents the unbound target, and AB still represents the aptamer-target complex:

$$A + B \overset{K_a}{\leftrightarrow} AB \tag{1.5.2.1}$$

From equation 1.5.2.1 and assuming a 1:1 ratio of aptamer to target in a complex, one can define the equilibrium association constant K_a as:

$$K_a = \frac{[AB]}{[A][B]} \tag{1.5.2.2}$$

With the additional assumption that the total amount of target B_T will always be equal to the sum of target in the complex and free target in solution:

$$[B_T] = [B] + [AB] \tag{1.5.2.3}$$

K_a is related to K_d as follows:

$$K_a = \frac{1}{K_d} \tag{1.5.2.4}$$

Equations 1.5.2.3 and 1.5.2.4 can be substituted back into equation 1.5.2.2 as:

$$K_a = \frac{1}{K_d} = \frac{[AB]}{[A]([B_T] - [AB])} \tag{1.5.2.5}$$

Rearranging gives a new expression for K_d:

$$K_d = \frac{[A]([B_T] - [AB])}{[AB]} \tag{1.5.2.6}$$

Dividing both sides by [A]:

$$\frac{K_d}{[A]} = \frac{[B_T]}{[AB]} - 1 \tag{1.5.2.7}$$

And rearranging once again gives:

$$\frac{[AB]}{[B_T]} = \frac{[A]}{K_d + [A]} \tag{1.5.2.8}$$

Equation 1.5.2.8 is a key form for the Michaelis-Menten equation.

At equilibrium and the amount of complex *[AB]* and total target B_T remain constant over time *t*:

Then: $$\frac{d[AB]}{dt} = 0 \, , \frac{d[B_T]}{dt} = 0 \quad (1.5.2.9)$$

And: $$\frac{[AB]}{[B_T]} = constant \quad (1.5.2.10)$$

Let this ratio of equation 1.5.2.10 be called the Michaelis-Menten constant *r*.

Thus: $$r = \frac{[A]}{K_d + [A]} \quad (1.5.2.11)$$

If we plot the measured value of *r* (or target complex over known total target) as a function of the measured value of free aptamer species at various aptamer concentration, one can generate a binding curve to to estimate the value of K_d. This method is straightforward, and has been used in current literature. [161, 162]

1.5.2.2 Methods to Determine Kd

Methods to measure K_d can be separated into two major categories: separation based methods and mixture based methods. This section will list the methods available to each category and very briefly summarizes some pros and cons for each based on a review by Jing and Bowser. [162]

1.5.2.2.1 Separation based methods

Separation based methods partition free aptamer from aptamer-target complexes to allow for their quantification. Separation based methods include: dialysis, nitrocellulose filter binding, ultrafiltration, gel electrophoresis (GE), capillary electrophoresis(CE), and high performance liquid chromatography (HPLC). Of these methods, CE and HPLC hold the potential for soluble targets such as proteins. They can also effectively be used as a partitioning tool during the SELEX process potentially while still measuring K_d of the aptamer selection pools; however, HPLC may require greater quantity of sample than CE. The other methods listed possess significant handling drawbacks. Dialysis requires a long time and possesses a low detection limit. Nitrocellulose filter binding requires radioactive labeling of sequences, and GE is sensitive but requires longer stretches of time.[162]

1.5.2.2.2 Mixture based methods

Mixture based methods attempt to quantify free aptamer and aptamer target-complex *in situ* or requires use in or after a separation process. Mixture based methods include: fluorescence, ultraviolet visible spectrophotometry (UV-Vis), circular dichroism (CD), surface plasmon resonance (SPR) and quartz crystal microbalance (QCM), and isothermal titration calorimetry (ITC). Of these, ITC gives the most information of all listed techniques about aptamer binding including ΔG, ΔH, ΔS, ΔC_p and K_d. QCM and SPR hold promise for evaluating immobilized aptamer without labeling schemes and ability to obtain thermodynamic and kinetic data.[163] Fluorescence labeling is widely used but can possibly interrupt aptamer structure and is influenced heavily by solution conditions. UV-

Vis lacks sensitivity and CD has not been widely used for K_d though it can provide conformational and stoichiometric data.

1.5.3 Motif Discovery

While there are many methods for evaluating K_d, determining K_d and other aptamer properties for many aptamer candidates can quickly become costly. For example, unlabelled DNA HPLC purification cost ~$100 a sequence for a 100 nMole synthesis and adding a FAM fluorescent tag for the same conditions increases the cost ~$150.[164] Following SELEX process, Sanger sequencing of individual colonies can return easily on the order of tens of sequences while bacterial plate Sanger sequencing can return hundreds of sequences, and NGS can return on the order of thousands of sequences. Thus, it is apparent that experimentally estimating K_d for several potential aptamer candidates can be cost prohibitive, not even considering additional costs associated with the target, immobilization and labelling chemistry, and equipment. Therefore, methods have continued to be developed and refined for picking aptamer candidates for further evaluation. Methods for determining promising aptamer candidates have stemmed from bioinformatics and computational biology techniques using sequence information (primary structure) and predicted secondary structure models of aptamers to identify motifs, or structural similarities and patterns that can predict affinity or function.

1.5.3.1 Primary Structure Motifs

Primary structure or sequence analysis of nucleic acids investigation was spurred by the advent of sequencing techniques and technologies which began to reach maturation in the mid-1960's. [165] With the ability to retrieve nucleic acid sequence information, databases

of sequences were created, and the need to develop methods for analysis became apparent.[166] In 1970, Needleman and Wunsch developed and algorithm to globally align or match overall similarity between two entire nucleic acid sequences through dynamic programming. [167] Later modification of this algorithm by Waterman and Smith allowed identification of stretches of multiple base segments of sequence similarity between two sequences through local alignment.[168] Finally, in the late 80s and early 90's, computational methods matured for microcomputers to allow simultaneous alignment of multiple sequences. [169, 170] Along with alignment methods, new ways of visually representing sequence patterns in DNA binding sites were developed known as a position specific scoring matrix (PSSM) [171] Lastly, multiple sequence alignments (MSAs) have been used to identify significant characteristics such as binding sites or local alignments shared by a group of sequences even if gaps (representing insertions, deletions, or substitutions, accumulated over, for example, the evolutionary pathways of two related species) exist in the alignment. This type of MSA is known as expectation maximization (EM).[172-174] EM has been combined with a mixture model to enable enhanced motif elicitation software known as Multiple EM for Motif Elicitation (MEME).[175] This section will briefly cover the use of these methods as tools for identifying candidate primary structure motifs in aptamer sequences.

Local alignment of sequences is undertaken through minimization of a scoring system of the matching between sequences. Local alignment is usually more meaningful than global alignments, as conserved patterns in sequences are included.[166] The local alignment scoring system can be described mathematically as between two sequences, sequence **a**=($a_1a_2 \ldots a_n$) and sequence **b**=($b_1b_2 \ldots b_n$) where $H_{ij} = H(a_1a_2 \ldots a_i, b_1b_2 \ldots b_j)$ then:

$$H_{ij} = max\{ H_{i-1,j-1} + s(a_i b_j), \max_{x \geq 1}(H_{i-x,j} - w_x), \max_{y \geq 1}(H_{i,j-y} - w_y), 0\} \quad 1.5.3.1.1$$

In which H_{ij} is the score at position *i* in sequence **a** and position *j* in sequence **b;** $s(a_i b_j)$ is the score for aligning the characters at position *i* and *j*; w_x is the penalty for inserting a gap base length of *x* in sequence **a**; and w_y is the penalty for inserting a gap base length of *y* in sequence **b**.[168] From equation 1.5.3.1.1, for alignment to be optimized, all S_{ij} pairings must be examined to determine a lowest score. As such, performing alignments on either long sequences or multiple sequences can be computationally demanding. Despite this computational challenge, these alignment tools are becoming increasingly used for aptamer sequence analysis.[59] An example of a local alignment is given in Figure 1.5.3.

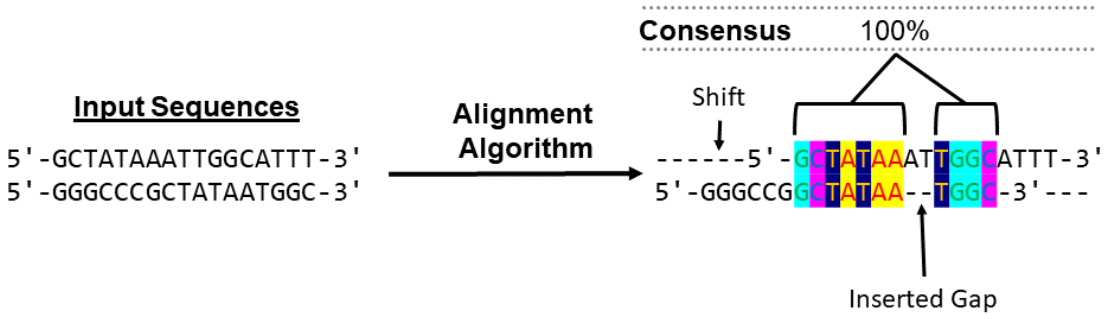


Figure 1.5.3. An example of a local alignment of two DNA sequences showing two regions of consensus due to inserted gaps.

Another technique to elicit information for potential binding motifs is the position specific scoring matrix (PSSM) or position weight matrix (PWM). [171] PSSM or PWM is performed by identifying the occurrence of a particular base identity and comparing it to a

relevant background mapping the frequency occurrence of a base identity at a particular position(to determine the "entropy" or uncertainty of these measurements) and then presenting the normalized mapping graphically in a base 2 or bit format. [166, 171] This method to display statistically significant base position sis increasing in popularity due to its visual effectiveness to describe these mathematical computations as illustrated in Figure 1.5.4.[59, 176]

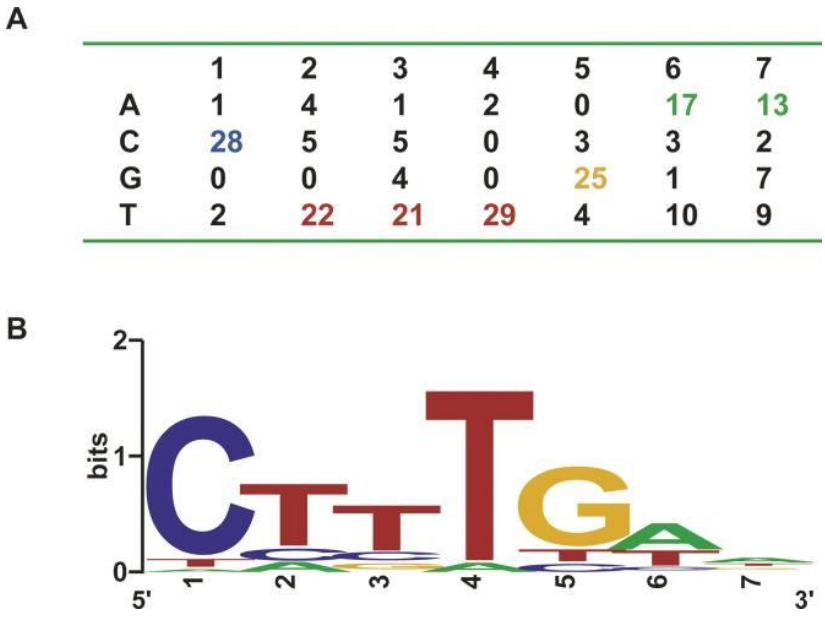


Figure 1.5.4 Illustration of (A) a matrix of base identity occurrences at each of 7 positions to calculate a PSSM and (B) the resultant graphical presentation of the PSSM showing larger bases at sites where they are likely more significant. The background presumes each of 4 bases has an equivalent frequency of 0.25 at each position. Taken from [177]

The final method for searching for the occurence of motifs is through expectation maximization(EM). Using the program MEME, it is possible to find motif segments, but without inserted gaps. [175] Since MEME is valid only for sequences with no gaps

inserted, it should be noted that EM methods identify motifs in a different manner from multiple sequence alignment tools, as shown in Figure 1.5.5.

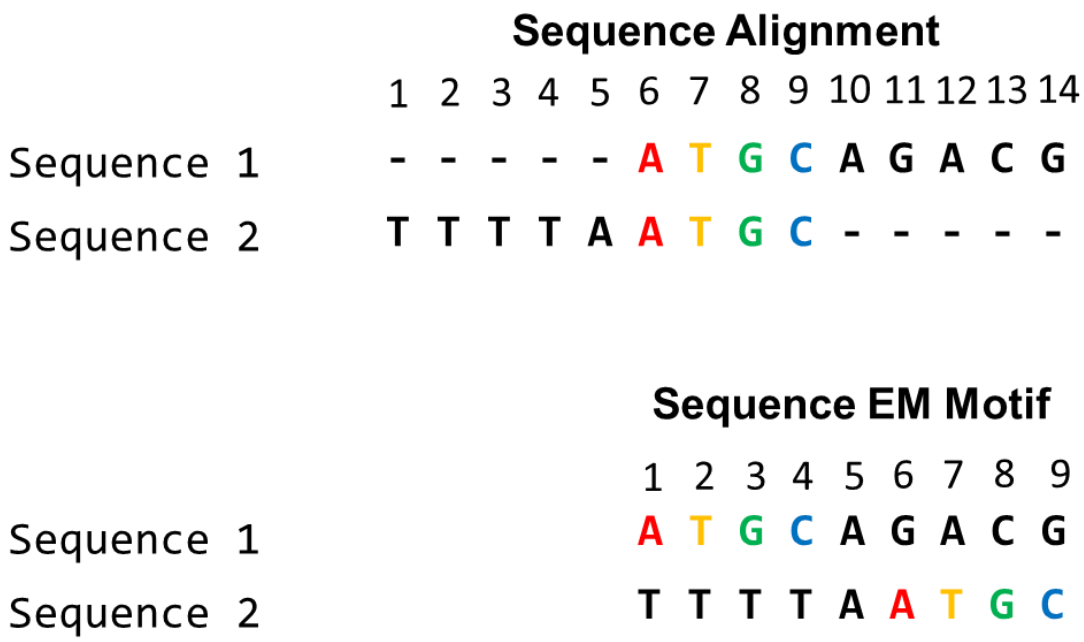


Figure 1.5.5. Differences in sequence alignment in which gaps are indicated with dashes and expectation maximization (EM) motif identification to identify base segment patterns

MEME analysis also outputs statistical significance values for an identified motif called the Poisson value or P-value and the expectation value or E-value. The P-value is the probability of the occurrence of the motif within just one sequence while the E-value is the ratio of the probability of the motif in the sequences scanned being unique to the probability of finding the identified motif in random sequences. Only motifs with an E-value<0.01 are considered significant to report. [172, 175]

1.5.3.2 Secondary Structure Motifs

It is generally accepted that the self-hybridized aptamer structure is relevant to target binding, and thus secondary structure motifs shared in a set of aptamers or aptamer candidates may point to target-binding motifs.[47, 178] Computational tools (e.g. UNAfold) provide predicted secondary structures and are heavily relied on as tools, especially since experimental verification of secondary structure is nontrivial. [15, 18, 20, 21] Notably, experimental studies are relatively rare, but selected studies have shown distinct differences in self-hybridization thermodynamics as compared to computational predictions.[179] On the other hand, predicted secondary structure of a VEGF aptamer was used to enable sequence modification to enhance the stabilization of aptamer structure as well as to increase affinity.[119] Others have attempted to create models for RNA aptamers that predict similarity based on secondary structures using previously reported aptamers sequences. [58, 59, 180-183] Bing *et al.* have demonstrated that aptamers selected for the same target by different groups exhibit comparable secondary structure motifs involved in binding. [47] Catherine *et al.* have used predicted secondary structure from an aptamer for thrombin to rationally design a structure switching aptamer for potassium ions. [179] Methods to identify secondary structure motifs are still being developed, but will likely become essential for creating "designer" libraries rather than simply relying on the diversity of randomized sequence libraries in next generation aptamer selection.

CHAPTER 2. NUCLEIC ACID ADDITIONS DURING SEED MEDIATED GROWTH OF GOLD NANORODS[122]

Note: This section was mainly taken from published work.

The effect of adding nucleic acids to gold seeds during the growth stage of gold nanorods (AuNR) is investigated using UV-Vis spectroscopy to reveal any oligonucleotide base or structure-specific effects (e.g. single-stranded versus double-stranded) on nanoparticle growth kinetics or plasmonic signatures. For seeds added to a gold nanorod growth solution, single-stranded polythymine induces a modest blue-shift in the longitudinal peak wavelength. Moreover, a particular sequence comprised of 50% thymine bases is found to induce a faster, more dramatic blue-shift in the longitudinal peak wavelength compared to any of the homopolymer incubation cases. In contrast to hompolymer additions, monomeric forms of the four nucleic acids, however, do not yield discernable spectral differences.

2.1 Introduction

Gold nanoparticles (AuNPs) have been heavily investigated for their unique size- and shape-dependent optical properties. These effects are apparent in the variation of optical spectra that gold nanoparticles exhibit for different sizes and shapes ranging from spheres (plasmon band at ~520 nm) to nanorods (transverse and longitudinal plasmon bands at ~520 and 600-1600 nm, respectively).[184-186] These signature spectral bands correspond to a localized surface plasmon resonance effect caused by the coherent, collective

oscillation of conduction band electrons in AuNPs illuminated with light.[184] In addition to the effects of size and shape on the resulting absorption and scattering of light, changes in the spatial arrangement of AuNPs can shift the position of the plasmon bands.[187-192] AuNP surfaces can be conveniently modified to incorporate other moieties such as oligonucleotides, proteins and antibodies due to the strong binding interaction between gold and chemical moieties such as thiols.[190, 193-196] As a result of these practical optical properties and conjugation possibilities, gold nanoparticles have been studied in biosensing, molecular imaging, therapeutic and medical diagnostic applications.[189, 195, 197]

Synthesis of gold nanoparticles of various shapes is typically performed via reduction of Au(III) ions in an aqueous solution with a capping agent present to stabilize formed particles. Common capping agents include citrate or cetyltrimethylamonium bromide (CTAB). [185, 194, 195, 198-201] To form gold nanorods (AuNR) gold seeds are typically added to a growth solution containing silver nitrate to promote preferential growth normal to the [202] facets of the gold nanorod in the [001] direction.[199] The aspect ratio of the AuNR can be controlled through choices in growth solution conditions such as changing the concentration of silver nitrate.[185, 198, 199, 203]

Over the past several years, numerous studies have focused on using grafted oligonucleotides to induce recognition-based assembly of gold nanoparticles as highlighted in recent reviews.[204-207] The studies typically conjugate single-stranded thiol-modified, complementary oligonucleotides on one or more populations of nanoparticles and then induce nanoparticle aggregation as hybridized duplexes bridge nanoparticle surfaces.[208-

210] As opposed to employing DNA as a nanoparticle assembly tool, select studies have explored the use of oligonucleotides and nucleotides as templates for the synthesis of inorganic nanoparticles. Similar to prior work using peptide-based templates[211-216], the goal of these bio-inspired approaches was to efficiently control particle shape and size during particle nucleation events, or more commonly, during the growth stage following seed preparation.[217] Some studies reveal, for example, that variation in physical size and fluorescence characteristics of CdS nanoparticles can be achieved by adding specific homopolymer sequences during nanoparticle synthesis.[218] DNA has also been used as a template to create Ag nanoring structures which could not be achieved using conventional direct growth methods.[219, 220]

Select studies by Wang *et al.* have specifically investigated the effect of incubating soluble oligonucleotides (i.e. not intentionally conjugated to material surface) with gold seeds during the growth stage. Their studies involved incubating solutions of 30 base-long (polyadenine, polythymine, or polycytosine)[221, 222] or 20 base-long (polyguanine)[44] homopolymers with CTAB-stabilized gold nanoprism seeds. They reported a variation occurs in particle morphology and topography ranging from rough (polyadenine) and smooth (polycytosine) surfaces to star-shaped (polythymine) and hexagonal (polyguanine) nanoparticles.[221] Depending on sequence compositions, binary mixtures of the homopolymers resulted in intermediate morphologies and surface roughness in the resulting nanoparticles.

While few studies[221, 222] have investigated particle growth in the presence of nucleic

acids, base-specific and structure-specific interactions of oligonucleotides with gold surfaces have been reported. SPR and FTIR analysis on gold surfaces can provide information on relative differences in the timing and extent of oligonucleotide adsorption events.[223-225] Notably, while bare planar gold serves as a convenient sample geometry for characterizing surface binding events, the nature of these binding events on colloidal gold may be significantly affected by additional materials parameters ranging from curvature to additional surface moieties (e.g. citrate).[224] For gold nanoparticles initially stabilized by citrate ions, bridging of nanoparticle surfaces via single-stranded homopolymers of DNA has been assessed using colorimetric analysis and UV-Vis spectroscopy.[226, 227] Collectively, these studies indicate that homopolymer sequences do not have equivalent affinities for gold substrates with purines adsorbing more strongly than pyrimidines to gold in the following order: G>A>C>T.[223, 226] Among these studies, the secondary structure of DNA (single-stranded chains vs. double-stranded helix) has been shown to affect the propensity for gold nanoparticles to aggregate.[225, 227]

In the current spectroscopy study, base-specific and structure-specific effects of single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) added during seeded growth are investigated to reveal if the presence of various nucleic acids directs the shape evolution of resulting gold nanoparticles. In an effort to directly compare results of adding various nucleic acids ranging from individual nucleotides to mixtures of polynucleotides in the current study, a common gold seed approach involving CTAB was chosen based on the work of Nikoobakht and El Sayed.[198, 199] While the seed preparation was identical in all cases, one of two growth conditions are used, namely either a gold nanosphere (AuNS) growth solution (aging of the original seed solution) or a gold nanorod (AuNR) growth

solution (seed solution added to a separate AuNR growth solution with additives such as silver nitrate). The effects of various nucleic acids on gold nanoparticle formation under these two broad classes of either nanosphere or nanorod growth solutions is primarily investigated using UV-Vis spectroscopy. Due to its widespread and practical use as a characterization tool for gold nanoparticle suspensions, UV-Vis spectroscopy was chosen as our primary analytical tool in order to quantitatively assess differences in the spectral signatures of the resulting nanoparticle suspensions from numerous nucleic acid incubation conditions. In order to identify base specific effects, 20 base-long homopolymer (**A20**, **T20**, **C20**, **G20**) and random (**R20**) sequences alone and as mixtures are employed in these studies. Structure-specific effects (ssDNA vs. dsDNA) were also investigated by incubating gold seeds with mixtures of complementary homopolymer sequences as well as with specific sequences (**S20**, **S20'**) that were shown to be capable of duplex formation (**S20**:**S20'**) under the nanoparticle growth conditions explored. Pure and mixed monomeric forms of nucleic acids were also investigated as additives during nanoparticle growth.

2.2 Gold Nanorod Specific Literature Review

As gold nanorods are used in this manuscript a brief review of the properties, biomedical applications, and synthesis methods is presented.

2.2.1 Properties and Biomedical Applications of AuNR

Gold nanorods possess an array of advantageous electronic properties that allow for diverse bio-based applications ranging from biosensing to therapeutics. As the size of the gold particle decreases from bulk to the nanometer scale, unique surface effects and quantum confinement dominate the optical properties.[198] These two sets of distinctive properties

of AuNR, radiative and non-radiative, and their biomedical applications are discussed in the next sections.

2.2.1.1 Radiative Properties

If excited by incident light, conduction band electrons of noble metals are subjected to a collective coherent oscillation in resonance with the frequency of light, a phenomenon referred to as surface plasmon resonance (SPR).[228] SPR causes the free electrons on the surface of the nanoparticle to undergo a dipolar oscillation which, in turn, affects the absorption of light depending on the dielectric properties, size, shape of the nanoparticles as well as the dielectric constant of the surrounding matrix material or solvent.[229] For anisotropic structures such as nanorods, different modes of oscillation are possible based on orientation, giving rise to shorter and longer characteristic wavelength absorption bands.[198] These absorption bands depend on the size and shape of the nanoparticle and facilitate tuning the optical response of the nanorods, even for a diverse population of nanoparticles as described by Gans Theory.[230, 231] After a plasmon has been created on a surface, however, it cannot continue to oscillate indefinitely due to the law of conservation of energy; thus, dephasing or dampening occurs through radiative (due to photon emission) and non-radiative (due to electron energy transitions and scattering resulting in heat dissipation) processes.[198, 232] The plasmon lifetime is inversely related to the linewidth at half maximum (FWHM) of the characteristic absorption band.[232] When SPR oscillation is generated on a nanoparticle, photons are restricted to the length of that nanoparticle (quantum confinement) which causes a decrease in plasmon wavelength (to fit the nanoparticle), but an increase in the amplitude of the photon wave which, in turn, significantly increases the electromagnetic field around the particle.[233]

This increase in electromagnetic field is termed the near-field effect and is responsible for the standard fluorescence activity of gold nanorods.[198] If these nanoparticles come into close proximity, interactions between particles occur due to coupling of their fields which will affect absorption bands of the particles.[234, 235] Another radiative property of AuNR with potential biomedical application is that of photoluminescence (PL). Though PL does occur in bulk gold, AuNR exhibit up to a six orders of magnitude increase in the quantum efficiency of the electron excitation and relaxation associated with PL.[236] These radiative properties enable AuNR to function as biosensors and molecular imaging aides. The SPR absorbance and scattering of AuNR have been exploited for application in plasmon sensing of antibodies [237, 238], protein [239, 240], and DNA [241, 242]. Coupled with surface-enhanced Raman scattering (SERS), AuNR have also been used to detect the presence of various chemicals[228, 243] and cancer cells[244-246]. SPR scattering of AuNR presents opportunities for detection and imaging techniques for *in vitro* [247, 248] as well as *in vivo* [249] conditions. Finally, the PL properties of AuNR have been coupled with two-photon luminescence (TPL) events to image cancer cells *in vitro* [250] and blood vessels *in vivo* [251].

2.2.1.2 Chemical and non-radiative properties

In addition to radiative properties, gold nanorods have other properties which facilitate their use in biomedical applications. Two well-known, non-radiative properties are the chemical (excluding catalytic properties) and physical properties of gold. As with a few other metals, sulfur atoms readily bind to gold surfaces enabling functionalization of the AuNR through thiol groups.[252, 253] In addition, oppositely-charged proteins or larger biomacromolecules can be adsorbed to the AuNR surface through electrostatic attraction

using pH adjustment[254, 255] and/or surface charge generation[248]. The chemical and physical surface properties of AuNR have been successfully utilized for surface modification through ligand exchange, electrostatic adsorption, surface coating, and use of biofunctional linkers to conjugate PEG, DNA, lipids, antibodies, various polymers, and other biologically-relevant molecules.[198, 256] AuNR also have a distinct set of photophysical properties applicable to biomedical purposes. When light is absorbed by the gold nanoparticle, the absorbed energy is converted into heat though photophysical events which are independent of size, shape, and plasmon mode.[189, 198, 257] These non-radiative events can heat the lattice faster than the surrounding matrix can cool the nanoparticle. If unchecked, this heating may result in reshaping and fragmentation of the particle, but the heating can be controlled through intensity of light exposure to prevent the photothermal melting of the nanoparticles. The photothermal behavior of AuNR has been used for photoacoustic tomographic imaging of cells [258](*in vitro* and *in vivo),* as a gene (and potential drug) delivery system[259], as well as an active and passive cancer cell targeting vehicle for subsequent photoirradiation of diseased cells[249, 260, 261].

2.2.2 Overview of Current Gold Nanorod Wet Synthesis Methods

The following section reviews current approaches in AuNR synthesis. Over time, synthesis approaches have evolved to provide high yields and better morphology control with faster, increasingly more facile methods. Gradual progress in nanoparticle synthesis typically resulted from "tweaking" of a particular method or use of specific reagents. An alternative approach relevant to the proposed work are biomineralization routes which offer a more biologically and environmentally friendly approach while remaining continuously facile. Here, the current colloidal AuNR synthesis methods are briefly highlighted.

2.2.2.1 Template Method

The earliest reported AuNR synthesis route is the template method.[262] This method involves a process in which gold is electrodeposited onto a porous polycarbonate or alumina substrate, which is then removed through chemical etching, leaving the AuNR in solution. The template method requires the use of strong acids and bases for the etching process and is limited by a relatively low yield. [198, 256, 263]

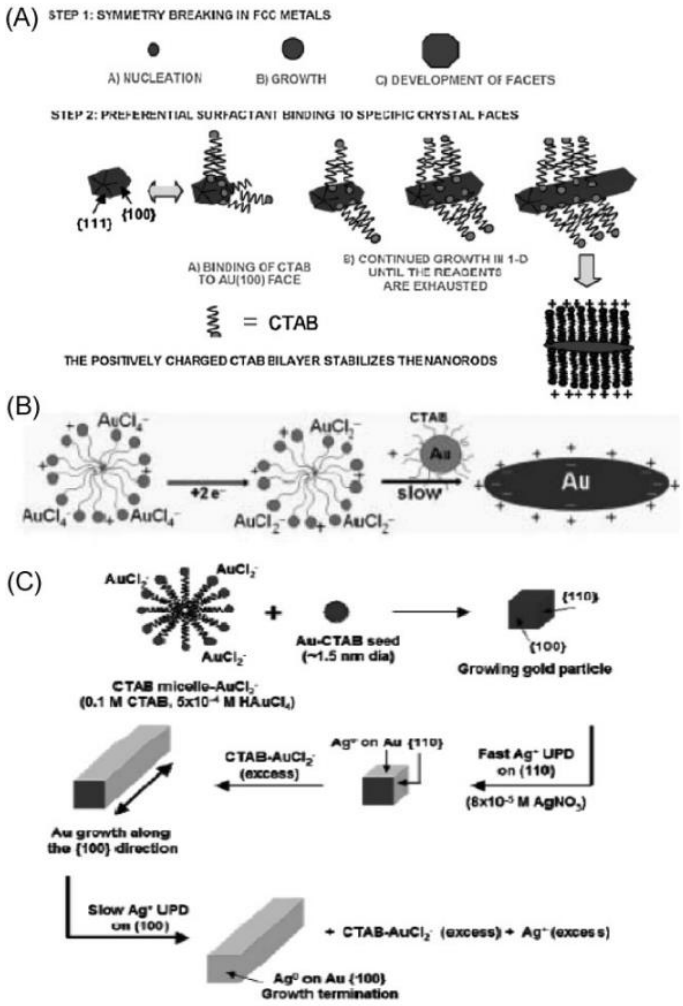


Figure 2.2.1 Mechanisms of seed-mediated growth of gold nanorods. (A) surfactant-preferential binding mechanism (B) electric field-directed mechanism. (C) underpotential deposition mechanism. Taken from[264]

2.2.2.2 Photochemical Method

The photochemical method is the first non-substrate-based synthesis of AuNR.[265] It involves a relatively simple light-sensitive reaction in which cationic micelles associated with auric acid under exposure to UV light (250-300) nm wavelength for up to 30 h. The UV light reduces the gold ions and initiates aggregation and growth in a single direction

promoted by the adsorbed surfactant.[198, 256, 266] Like many of the following methods, common reagents used can include cetyl trimethyl ammonium bromide(CTAB), silver nitrate, acetone, and cyclohexane.

2.2.2.3 Electrochemical Method

Due to its high yield, the electrochemical method is a popular approach.[198] In this method, AuNR are formed by the electrolysis of a gold anode and platinum cathode associated with a silver plate in a solution of CTAB, tetraoctylammonium bromide(TOAB), acetone, and cyclohexane.[267] During electrolysis, the gold ions are bound to the bromine ions and associate with the surfactant micelles (as illustrated in Figure 2.2.1(B) which migrate to the cathode. At the cathode, the gold ions are reduced and resulting in AuNR deposition as illustrated in Figure 2.2.2 (C). [198, 256, 268]

2.2.2.4 Seeded growth method

The most popular AuNR synthesis, the seeded growth method, offers yields of ~99% within 20 min. This method results in AuNR with tunable aspect ratios from 1.5 up to 70 through control of growth solution conditions.[198, 199] A three step procedure developed by Jana et al.[269] was later improved by Nikoobakht and El-Sayed to achieve high yield and monodisperse rods such as the one shown in Figure 2.2.2(B) [199] A seed solution is prepared by reducing auric acid in a CTAB solution by ice-cold sodium borohydride. The seed solution is then added to a growth solution (CTAB, auric acid, ascorbic acid, and an adjustable amount of silver nitrate for tunable aspect ratio) as depicted in Figure 2.2.1(C). For higher aspect ratios of the AuNR, a benzyldimethylalkylammonium chloride(BDAC) strong co-surfactant is added to the growth solution.[199, 256]

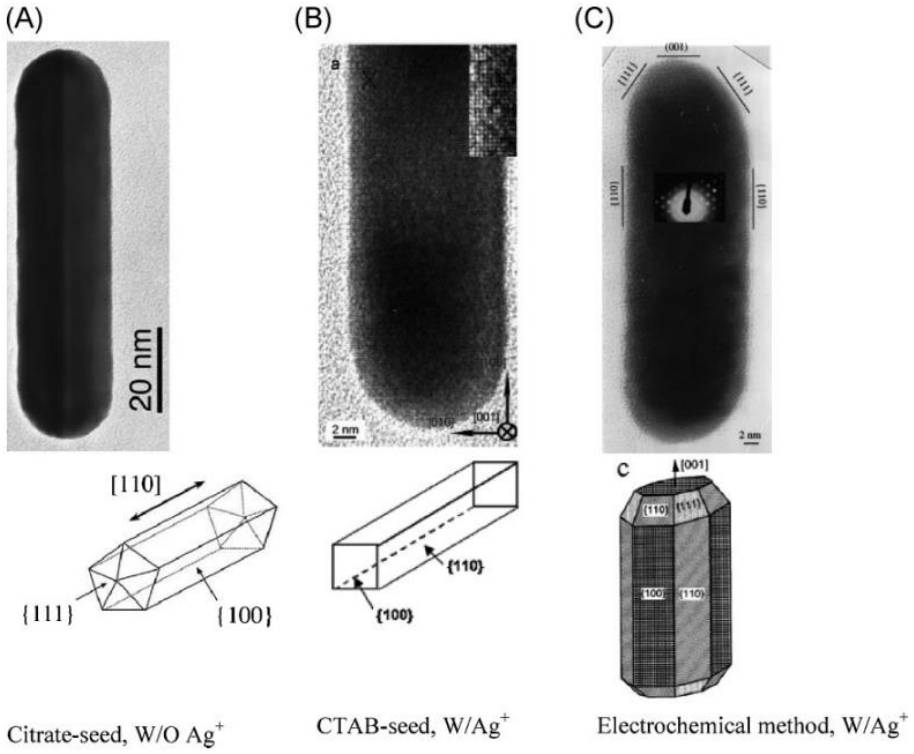


Figure 2.2.2. AuNR morphologies produced by: (A) Seed-mediated growth with citrate cap and Ag+; (B) Seed-mediated growth with CTAB cap and Ag+; (C) Electrochemical method with Ag+. Taken from [198]

2.2.2.5 Seedless growth method

More recent studies have avoided the preparation of a seed solution in favor of flow-based production of AuNR.[270] For this one step method, sodium borate is added directly to the same growth solution used in the seeded method. While high yield and monodispersity are obtained with this method, the range of aspect ratio is limited.[271] The growth kinetics are also slow at room temperature which has motivated more recent attempts to use temperature conditions approaching 96 ˚C to achieve faster nanoparticle formation.[46, 47]

2.2.2.6 Biomineralization approaches

Previous attempts to form gold nanoparticles (AuNP) through biomineralization pathways have shown promise though the rod morphology has not prevailed in the resulting nanoparticle population. Various routes have been explored for AuNP synthesis in the presence of biological species including gold ion reduction with live fungus[214], alfalfa plant[215], or bacteria[213]; bio-reduction with *Aloe Vera* extract[272], alfalfa extract[273], or lemongrass extract[274]. Another approach involves panning for specific peptide sequences that can precipitate particular AuNP shapes based on peptide sequence[212]. While the range of AuNR morphology has been limited thus far, these successful results with peptides suggest that bio-macromolecules may be useful for AuNR synthesis, especially if shape can be readily controlled. The Lu Group at University of Illinois, Urbana-Champaign has recently investigated how individual DNA bases affect the shape and even the surface roughness using prism-shaped gold seeds as shown in Figure 2.2.3. [221, 222]

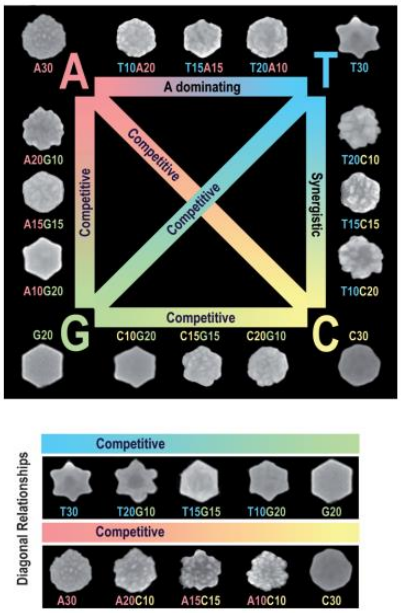


Figure 2.2.3 Mapping of effect of adding oligonucleotides at various base compositions on gold nanoparticle morphologies during seeded growth of nanoprism seeds. Taken from [243]

2.3 Materials and Methods

2.3.1 Materials

Chemicals. Hydrogen tetrachloroaurate (III) hydrate ($HAuCl_4 \bullet 3H_2O$), silver nitrate ($AgNO_3$), L-ascorbic acid, hexadecyltrimethylammonium bromide (CTAB), hydrochloric acid (HCl), and deoxynucleoside monophosphates (dNMPs: dAMP, dTMP, dGMP, and dCMP) were purchased from Sigma Aldrich, St. Louis, MO. Sodium borohydride ($NaBH_4$) was purchased from Fluka Analytical, Bushs, Switzerland. Tris HCl (pH 7.5) was purchased from Amresco, Solon, OH. Deoxynucleoside triphosphates (dNTPs: dATP, dTTP, dGTP, dCTP, dNTP Mix) were purchased from Invitrogen, Grand Island, NY. All

DNA sequences were synthesized by Integrated DNA Technologies, Inc., Coralville, IA. Buffers were prepared using 0.2 µM filtered water from a Barnstead Nanopure ultrapure water purification system. 1 mM stock solutions of DNA were prepared using 1 mM Tris HCl, pH 7.5.

2.3.2 Preparation of Gold Nanoparticle Seed Solution and Gold Nanorods

Preparation of gold nanoparticle seed solution. Gold nanoparticle seeds were synthesized according to the methods described by Nikoobakht and El Sayed.[199] CTAB solution (20 mL, 0.2 M) was mixed with a solution of $HAuCl_4 \bullet 3H_20$ (20 mL, 0.5 mM). After 30 min of stirring, a freshly prepared solution of ice-cold $NaBH_4$ (2.4 mL, 0.01 M) was then added to the stirred seed solution and mixed for an additional 2 min. Gold seed suspensions were then added to nucleic acid solutions as described in the next section.

Preparation of gold nanospheres and nanorods. Gold nanoparticle seeds for gold nanorod synthesis were prepared at the reagent ratios as described above; the total reaction volume (including additional $NaBH_4$) was scaled to 10.6 mL. Using magnetic stir bars, the seed solution was then stirred on a low setting for an additional 2 min before being poured into a separate glass container and set aside for 45 min to 1 hr prior to incubation with (or without) various nucleic acids in a gold nanorod growth solution described next.

A 10 mL gold nanorod growth solution was prepared to provide the same growth conditions for each set of nucleic acids for a given nanoparticle suspension (1 mL each). For cases involving a large number of samples (e.g. kinetics studies over a 7 day timeframe), the total reaction volume was scaled accordingly. Then, for each 10 mL volume, 160 µL of 5 mM $AgNO_3$ was added to 5 mL of 0.2 M CTAB and stirred on a low

setting for 30 s. A 5 mL volume of 1 mM $HAuCl_4 \cdot 3H_20$ was added to the solution and stirred on a low setting for 5 min. A 54 µL volume of 0.1 M L-ascorbic acid was then added to the solution and stirred on a low setting for 5-10 min. To then initiate AuNR growth, 12 µL of a gold seed solution (aged 45 min to 1 h) was added to 10 mL of AuNR growth solution. The solution was immediately stirred on a high setting for 15-20 s and then added to various nucleic acid solutions described below.

2.3.3 Incubation of Nucleic Acids with Gold Nanoparticle Seeds in Gold Nanorod Growth Solution

A list of oligonucleotide sequences is provided in . For all seed suspensions incubated with nucleic acids, each oligonucleotide or nucleotide was incubated at a final concentration of 2 µM in a 1 mL total volume of the seed solution added to the AuNR growth solution. For cases involving mixtures of oligonucleotides or nucleotides, each oligonucleotide or nucleotide was added to yield a final concentration of 2 µM for each individual species. Select experiments were also conducted at a nucleic acid concentration of 0.1 µM. Following the addition of seed solutions to a particular nucleic acid solution, suspensions were mixed on a rotomixer and examined at various timepoints using UV-Vis analysis. Select samples were additionally characterized via transmission electron microscopy (TEM).

2.3.4 UV-Vis Spectroscopy and TEM Analysis

For the early time points in these particular kinetics studies, UV-Vis spectra was collected with a sweeping scan at a 1 nm wavelength resolution. For all other time points occurring at or after 1 h, however, the previously described scanning conditions (i.e. 1 nm intervals)

was used. After data collection, spectra for each set of samples were first normalized against a water blank. Next, each sample was normalized to the same relative absorbance value of 0.3 at 400 nm as others have done in previous studies.[275] Select samples were additionally characterized via transmission electron microscopy. TEM images were acquired with a JEOL 100CX II transmission electron microscope at 100 kV. Samples for TEM were centrifuged (14,000 rpm x 15 min) twice to remove surfactant and redispersed in nanopure water. After redispersion, 0.3 μL of the sample was drop-cast onto a carbon-coated copper TEM grid and allowed to dry at room temperature.

Table 2.3.1. List of 20 base-long DNA sequences incubated with gold seeds in the presence of AuNR growth solution.

Nomenclature	Sequence (5'→3')
A20	AAA AAA AAA AAA AAA AAA AA
T20	TTT TTT TTT TTT TTT TTT TT
C20	CCC CCC CCC CCC CCC CCC CC
G20	GGG GGG GGG GGG GGG GGG GG
R20	25% (A,T,C,G) in random order
S20	TTT TTT GGA TTG CGG CTG AT
S20'	ATC AGC CGC AAT CCA AAA AA

2.4 Results and Discussion

2.4.1 Effects of 20 base-long homopolymer additions to gold seeds in gold nanorod growth solution

In the next series of DNA incubation studies, freshly-prepared (within 45 min - 1 h) gold nanoparticle seeds are first added to a gold nanorod (AuNR) growth solution and this mixture is then added to various nucleic acid solutions. Following a 2 h incubation time, two peaks are observed to form in either the absence (controls) or presence of DNA as shown in Figure 2.4.1(a). Prior studies indicate that the dual peaks are associated with the transverse and longitudinal axes of AuNR.[185, 198] Differences between the transverse peak wavelength values (~520 nm) are typically small; however, differences in the longitudinal peak wavelength values (~600-1600 nm) between various AuNR suspensions

can be significant and are generally attributed to differences in the aspect ratio of the AuNR (e.g. a lower longitudinal peak wavelength value corresponds to nanorods with a lower aspect ratio[276]).

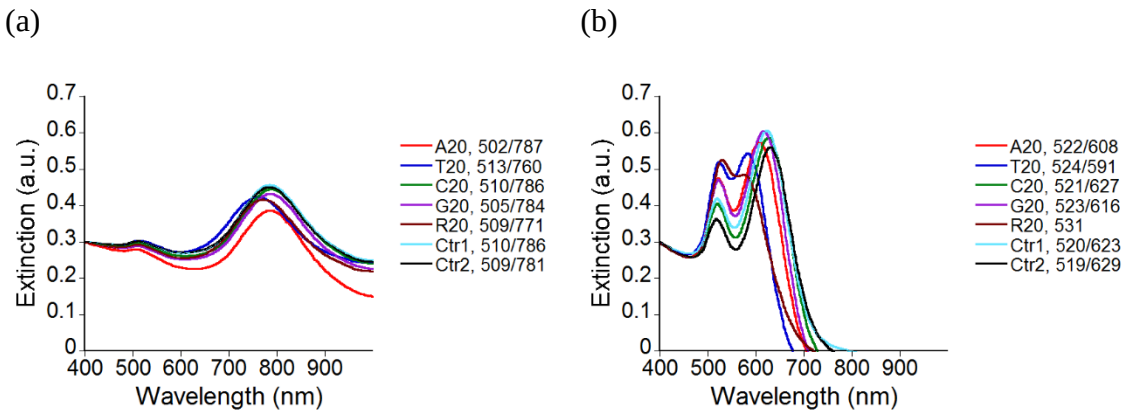


Figure 2.4.1. UV–Vis spectra of gold seed in AuNR growth solution following incubation with various 20 base-long homopolymers (A20, T20, C20, G20) and random (R20) sequences for 2 h (a) and 7 d (b). Controls involve the addition of 2 µL of Tris HCl (Ctr1) or 2 µL of nanopure water (Ctr2) in the absence of DNA. The resulting peak wave length values are included in the legend

As compared to the control cases, the most significant spectral differences at the 2 h timepoint involve a modest blue shift in the presence of polythymine, **T20**, and the 20 random base DNA, **R20**, cases. At a lower DNA concentration of 0.1 µM, however, spectral differences are not as apparent for any of these sequences as shown in Figure 2.4.2(a). As these suspensions continue to age in AuNR growth solution for 7 d, the longitudinal peak undergoes a blue shift in every case, but to a different extent depending on the sequence as shown in Figure 2.4.1(b). In previous reports this shift to lower longitudinal peak wavelength values with prolonged growth time in the absence of DNA has been attributed to nanoparticle overgrowth resulting in nanorod-like shapes with smaller aspect ratios.[277] In particular, the random **R20** sequences appear to ultimately

have the most dramatic effect on this blue-shift effect of the longitudinal peak. In fact, while transverse and longitudinal peaks are still evident by the seventh day in the controls and homopolymer cases, these peaks nearly merge in the **R20** case, even at a lower concentration of 0.1 μM (Figure 2.4.2(b)). Of all the homopolymers, it is polythymine, **T20**, that has the most dramatic effect on peak shifts over time as shown in Figure 2.4.1(b).

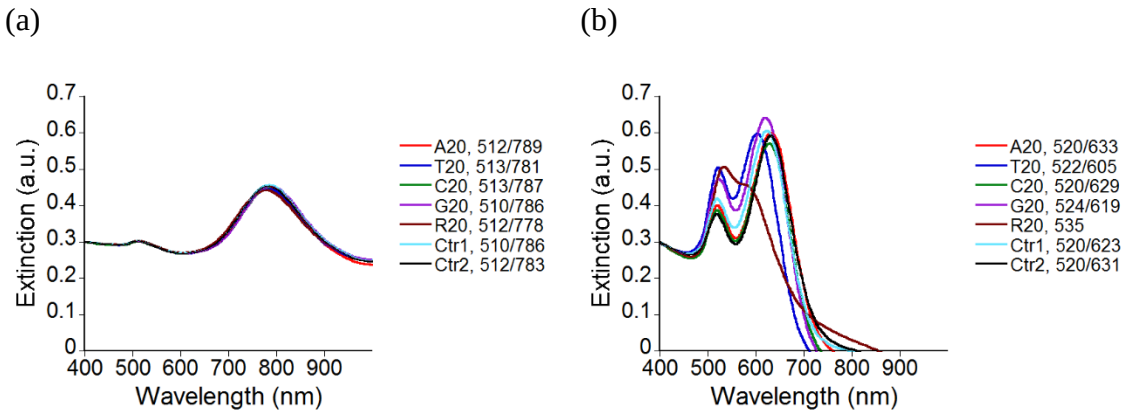


Figure 2.4.2. UV-Vis spectra of gold seed in AuNR growth solution following incubation with various 20 base-long homopolymers (A20, T20, C20, G20) and random (R20) sequences at 0.1 μM for 2 h (a) and 7 d (b). Controls involve the addition of Tris HCl (Ctr1= 2 μL Tris HCl) or nanopure water (Ctr2= 2 μL 18 MΩ-cm water) in the absence of DNA. The resulting peak wavelength values are included in the legend.

AuNR growth solution with dNTPs do not result in any significant differences in the dual peaks observed in the UV-Vis spectra as shown in Figure 2.4.3. Thus, this result further indicates that the polymeric nature of the oligonucleotides plays a key role in the spectral evolution of these nanoparticle suspensions.

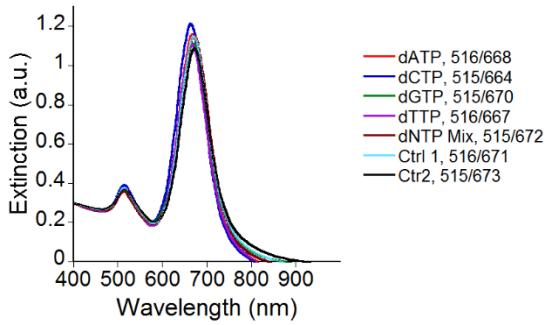


Figure 2.4.3. UV-Vis spectra of gold seed in AuNR growth solution following incubation with various dNTPs (dATP, dTTP, dCTP, dGTP, dNTP Mix); Controls involve the addition of Tris HCl (Ctr1= 2 µL Tris HCl) or water (Ctr2=2 µL 18 MΩ-cm water) in the absence of DNA. The resulting peak wavelength values are included in the legend.

2.4.2 Effects of homopolymer mixture additions to gold seeds in gold nanorod growth solution

Given the fast evolution of peak formation and the shifts in the longitudinal peak wavelength values over time for AuNR growth solution conditions shown in Figure 2.4.1 (a,b) experiments with homopolymer mixtures in AuNR growth conditions were conducted to evaluate the effects of the **T20** sequence on resultant spectra after a 2 h incubation (instead of 7 d).

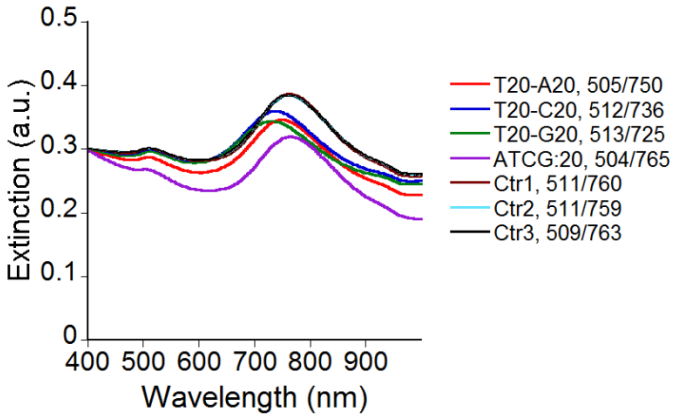


Figure 2.4.4. UV–Vis spectra of gold seed in AuNR growth solution after a 2 h incubation in the presence of various homopolymer mixtures of two or four sequences. Controls involve the addition of Tris HCl (Ctr1 4 µL Tris HCl; Ctr2 8 µL Tris HCl) or water (Ctr3

4 µL nanopure water) in the absence of DNA. The resulting peak wave length values are included in the legend.

Figure 2.4.4 shows that blue-shift resulting from the presence of **T20** is increasingly suppressed by the copresence of one other homopolymer in the following order: **G20** < **C20** < **A20** as well as by the presence of all four homopolymers. It is likely that specific base effects of single-stranded polythymine on seed growth in AuNR growth solution is inhibited by Watson-Crick base pair matching in the **T20**-**A20** mixtures.

2.4.3 Effects of 20 base long complementary strands alone and mixed together on gold seeds in gold nanorod growth solution

In contrast to the early appearance of a single peak in aged gold seed solutions incubated with **S20:S20'** duplexes as described above, gold seed in AuNR growth solution incubated with the **S20:S20'** duplex solutions exhibits a modest blue shift of the longitudinal peak compared to the controls at the 2 h time point as shown in Figure 2.4.5(a).

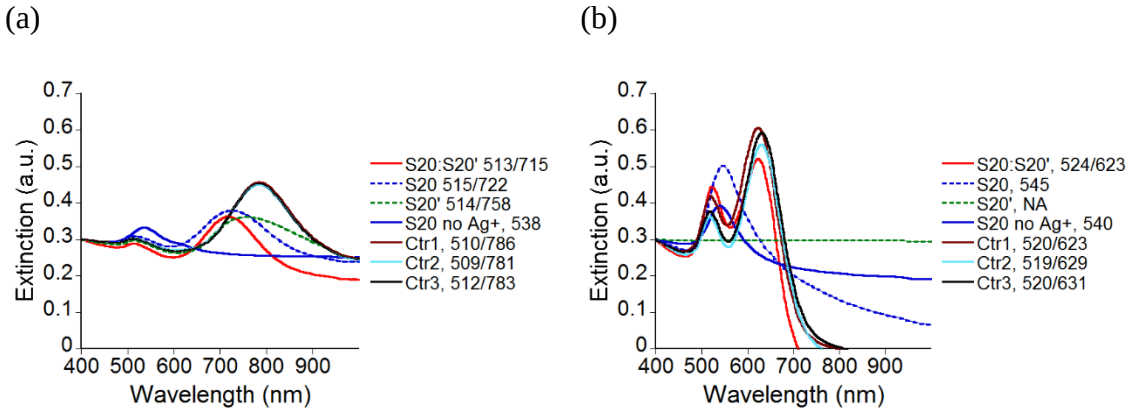


Figure 2.4.5. UV–Vis spectra of (a,b) gold seeds and (c,d) gold seeds in AuNR growth solution following incubation with various complementary ssDNA alone (S20 or S20') and mixed together (S20:S20') for 2 h (a) and 7 d (b). Controls involve the addition of Tris HCl (Ctr1 4 µL Tris HCl; Ctr2 2 µL Tris HCl) or water (Ctr3 2 µL nanopure water) in the absence of DNA. The resulting peak wave length values are included in the legend. The

letters “NA” in this legend for S20' in (d) indicate that the suspension spectra could not be measured due to formation of an aggregated pellet that could not be redispersed on day 7.

At the 7 d timepoint, however, there was little difference in the translational and longitudinal peak wavelengths (~520 nm and ~620 nm) between the **S20:S20'** case and the control samples with no DNA additions as shown in Figure 2.4.5 3(b). The thymine-rich **S20** case, on the other hand, promotes the most significant effect on the spectra causing a large blue-shift of the longitudinal peak at the 2 h time point which continues to shift until it ultimately merges with the translational peak to form a single peak at 545 nm the 7 d timepoint. Intriguingly, this blue-shift effect for this thymine-rich (50%) **S20** sequence is more dramatic than that observed for the pure polythymine case shown in Figure 2.4.53(a,b). Moreover, the adenine-rich (50%) **S20′** sequence causes only a modest blue-shift during the early incubation time in Figure 2.4.5(a), but causes extensive aggregation over the long incubation time shown in Figure 2.4.5(b). The featureless spectra shown in Figure 2.4.5(b) for the **S20′** case is due to the fact that the aggregate ultimately adheres to the reaction vessel wall during the 7 d incubation time and could not be redispersed for spectroscopic evaluation. In contrast to the **S20** and **S20′** cases, the longitudinal peak undergoes a more modest blue-shift in the other cases shown in Figure 2.4.5(a,b) in which two peaks remain evident at all timepoints. The presence of the single peak by day 7 in the **S20** case suggests that the particles are likely to be more spherical in shape as indicated by TEM micrographs shown in Figure 2.4.6(a).

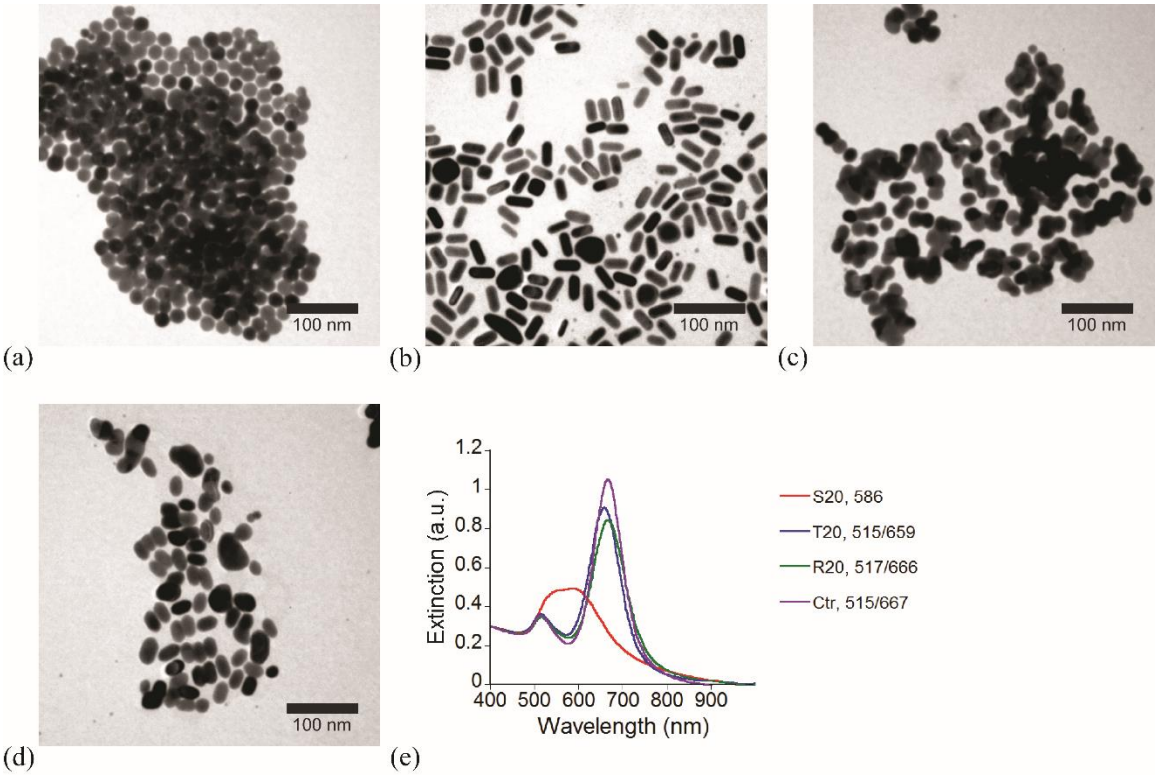


Figure 2.4.6 TEM micrographs of 2 µM incubation samples after 3 days (a) S20 (b) T20 (c) R20 (d) Control= 2 µL Tris HCl and (e) accompanying UV-Vis spectra of all incubation samples. The resulting peak wavelength values are included in the legend. (TEM performed by Maeling Tapp).

As a separate set of experiments, additional incubation studies conducted in the absence of silver nitrate indicate only a single peak occurs at the 2 h and 7 d timepoints in either the presence (see Figure 2.4.5(a,b)) or absence (results not shown) of **S20** strands. Silver ions reportedly promote the anisotropic growth of the CTAB-stabilized gold seeds into gold nanorods through its preferential deposition onto {110} faces during particle growth, effectively inhibiting growth on this plane and promoting growth in the [100] direction on {111} facets. [264] While the formation of nanorods themselves does require the presence of silver nitrate, the presence of DNA sequences such as the thymine-rich **S20** does appear to significantly affect the spectral evolution of gold nanorods and may even

inhibit the role of silver nitrate in promoting stable gold nanorod growth over time. A lower concentration of **S20** (0.1 µM) does not promote significant differences in the AuNR spectra compared to the controls as shown in Figure 2.4.7(a,b).

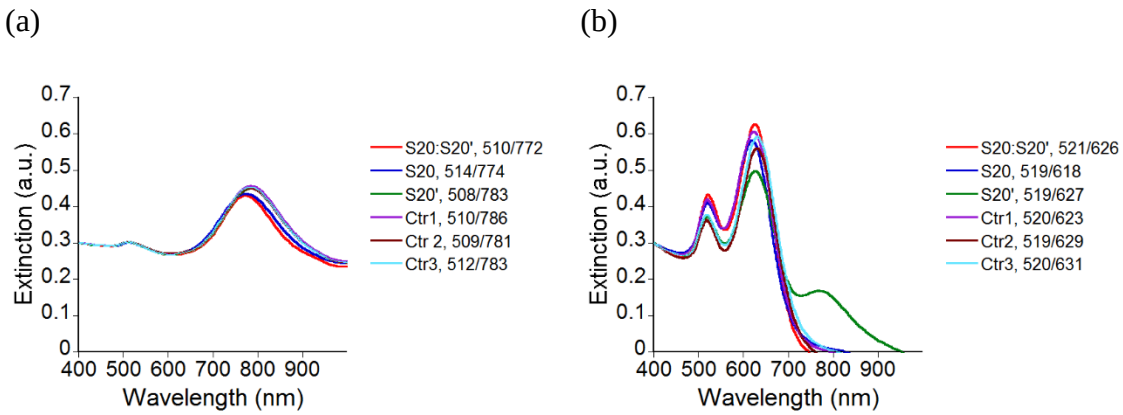


Figure 2.4.7. UV-Vis spectra of gold seeds in gold nanorod growth solution following incubation with various complementary ssDNA alone (S20 or S20') and mixed together (S20:S20') at 0.1 µM for 2 h (left) and 7 d (right). Controls involve the addition of Tris HCl (Ctr1= 4 µL Tris HCl, Ctr2= 2 µL Tris HCl) or water (Ctr3= 4 µL 18 MΩ-cm water) in the absence of DNA. The resulting peak wavelength values are included in the legend.

2.4.4 Effects of select ssDNA sequences on spectral evolution of gold nanorod growth solutions

The ssDNA cases (**T20**, **R20**, and **S20**) promoting the most dramatic blue shift in the longitudinal peak wavelengths from studies described above were selected for a more detailed study to monitor peak evolution over several timepoints during the first 2 h (Figure 2.4.8) and first 7 d (Figure 2.4.9). As shown in Figure 2.4.8(a-d), the characteristic transverse and longitudinal peaks appear within 15 min of adding seeds to AuNR growth solution in either the absence (controls) or presence of DNA.

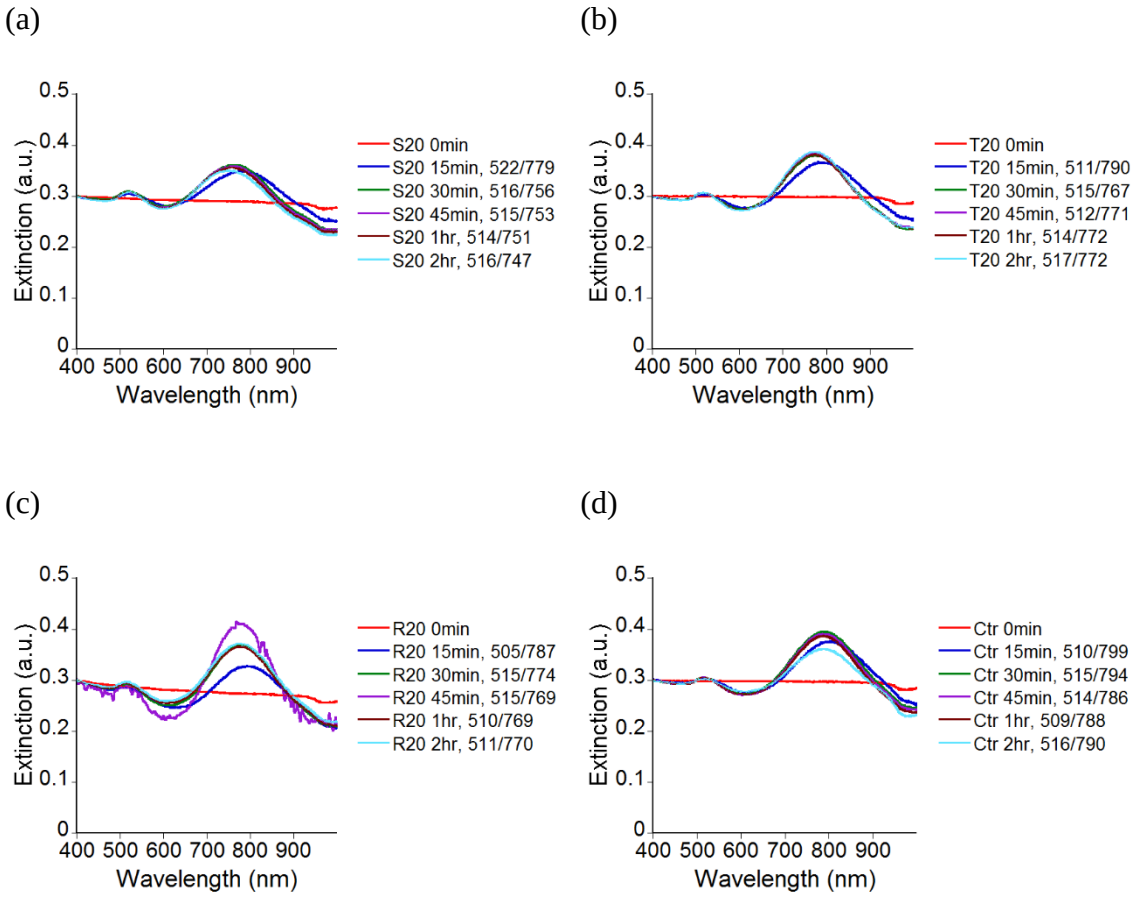


Figure 2.4.8. UV–Vis spectroscopy over the first 2 h (in 15 min intervals) of gold seeds in AuNR growth solution incubated with the following solution additions: (a) S20, (b) T20, (c) R20, or (d) 2 µL Tris HCl only (no DNA). The resulting peak wave length values are included in the legend.

As reported previously by Zweifel *et al.*[277] a blue shift of the longitudinal peak of AuNR during growth does occur within the first 2 h in the absence of DNA; however, for the control case shown in Figure 2.4.8(d), only a small shift ($\Delta\lambda$ = 9 nm) in the longitudinal peak location occurs over this 2 h period. Larger shifts, on the other hand, are observed in the presence of DNA for the **S20** ($\Delta\lambda$ = 32 nm), **T20** ($\Delta\lambda$= 28 nm) and **R20** ($\Delta\lambda$= 17 nm) cases. These suspensions involving **S20**, **T20**, or **R20** additions were then monitored daily over a 7 d period as shown in Figure 2.4.9(a-d).

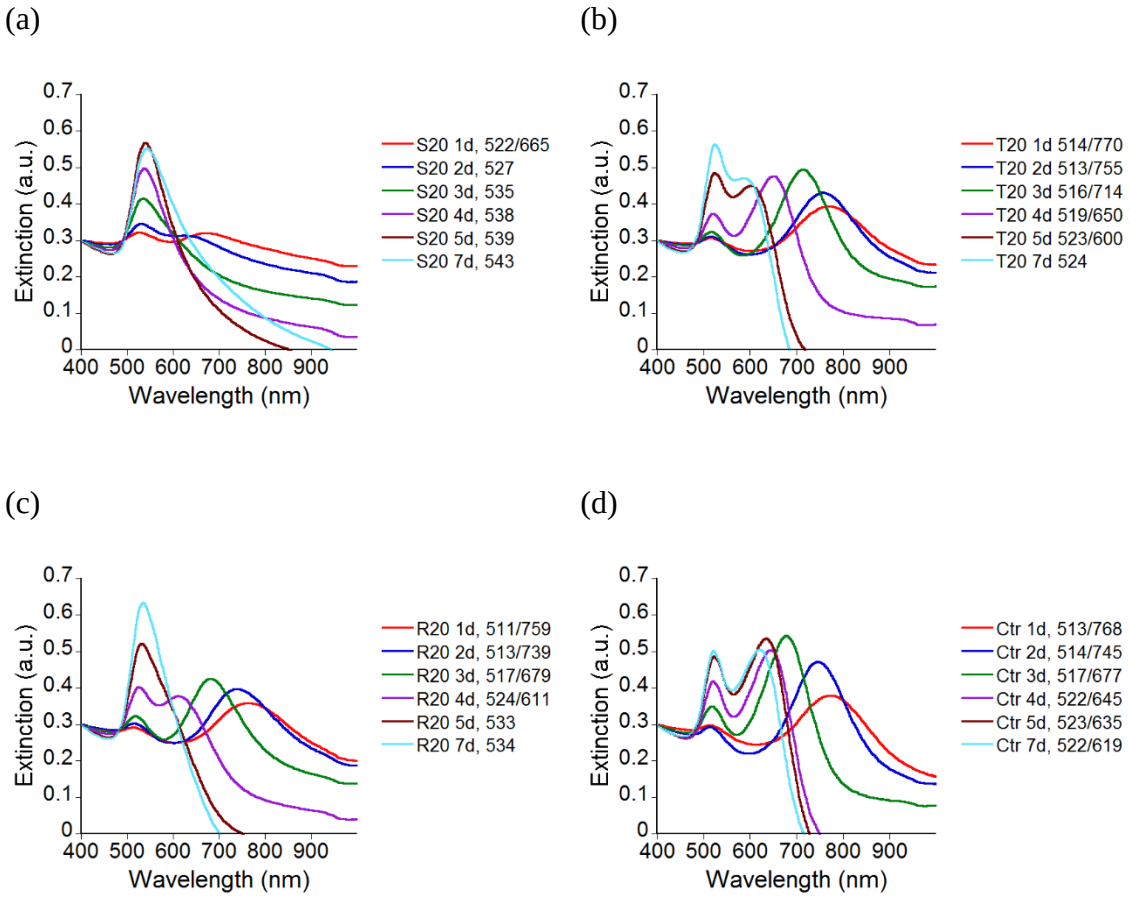


Figure 2.4.9. UV–Vis spectroscopy over the first 7 d (1-day intervals) of gold seeds in AuNR growth solution incubated with the following solution additions: (a) S20, (b) T20, (c) R20, or (d) 2 µL Tris HCl only (no DNA). The resulting peak wave length values are included in the legend.

Within 1 d, **S20** induces the most dramatic blue shift in its longitudinal peak wavelength ($\Delta\lambda$ = 114 nm) as compared to its 15 min time point value. The dramatic blue shift continues for the **S20** case until this longitudinal peak merges with the translational peak to form a single peak (535 nm) within 3 days. A similar merger of peaks (533 nm) occurs by day 5 for the **R20** case. While a blue shift of the longitudinal peaks continues for the **T20** and control cases, two distinct peaks still remain apparent by day 7 for these two cases. For more facile comparison, the relative change in the longitudinal peak wavelength

value over time is plotted in Figure 2.4.10 using spectra presented in Figure 2.4.8 and Figure 2.4.9 for the **S20**, **R20**, **T20**, and control cases. As shown in Figure 2.4.10, larger blue-shifts in the longitudinal peak wavelength values are generally observed in the presence of any of these DNA sequences than in the absence of DNA, particularly at later time points.

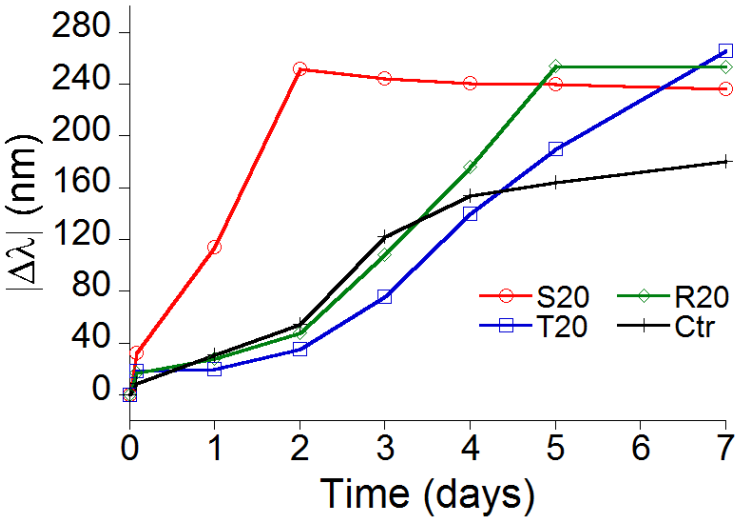


Figure 2.4.10. Relative difference between AuNR longitudinal peak values over time relative to the first observed peak timepoint (at 15 min) for various DNA sequence incubation conditions.

Of these DNA sequences, the shift in peak values occurs most rapidly for the **S20** case. In contrast to the shifts observed for the longitudinal peak, the transverse peak (ranging from 505 to 530 nm, depending on suspension conditions) remains relatively unchanged over time for a given suspension. This result is not surprising since the transverse peak is reportedly relatively insensitive to differences in the aspect ratio of gold nanorods.[230] Additionally, it has been shown that suspensions with dispersed gold nanospheres with a size range of 5-20 nm have a similar characteristic peak value (~520 nm).

In addition to the spectral analysis in which nanoparticle shape information can be inferred, TEM was performed following a 3 d incubation with these select sequences as shown in Figure 2.4.6. As expected for the **S20** incubation case in which only a single UV-Vis peak was observed for the nanoparticles, only spherically-shaped nanoparticles are found (see Figure 2.4.6(a,e)). The **T20** incubation case yields mostly nanorods with a few larger spherical and oblong particles (see Figure 2.4.6(b)) while **R20** and the control (no DNA) resulted primarily in oblong nanoparticles with a small aspect ratio (Figure 2.4.6(c) and (d)).

2.5 Conclusions

In the current work, monomeric forms of the nucleic acids do not appear to affect the spectral evolution of growing nanoparticles. In contrast, base-specific and structure-specific effects of 20 base-long oligonucleotides are evident in several cases involving seed-mediated growth of gold nanoparticles. Moreover, while the distinctive and likely dynamic role of the cationic CTAB species is not clear, separate washing studies indicate that CTAB is only weakly bound to the gold surface. Thus, displacement by a stronger adsorbate to cause aggregation is possible. Among the homopolymers, modest base-specific effects are observed in terms of the timing and extent of the blue-shift of the longitudinal peak of the gold nanorods in the case of polythymine. However, the copresence of the complementary homopolymer lessens the effect of the homopolymer alone. Collectively, the interactions between complementary oligonucleotides involving Watson-Crick base pair matching appears to either completely overcome (for blunt-ended duplexes) or at least reduce (for duplexes with single-stranded overhangs) interactions between oligonucleotides and nano-sized gold particles during their growth

CHAPTER 3. COMPELS SCREENING TO IDENTIFY DNA APTAMERS FOR GOLD NANORODS

In this chapter, a non-evolutionary screening process known as Competition Enhanced Ligand Selection (CompELS) is employed to screening a large (~10^{16}) random library of 69 base-long, single-stranded DNA sequences to identify several (42) aptamer candidates for gold nanorods. Identified aptamers are examined for similarity in primary structure with multiple sequence alignment (MSA) and multiple EM for motif elicitation (MEME). Primary structure analysis does not reveal any clear patterns in individual bases at specific position, nor any clear patterns in base segments within the 40 base-long central segment of the aptamer sequences. This lack of pattern in bases is not surprising since the CompELS screening process itself does not inherently promote an evolutionary screening pathway.

3.1 Introduction

SELEX is an iterative screening process comprised of the following key steps during each selection round: (1) incubation of a large, random sequence library with the target of interest; (2) removal of unbound or weakly bound sequences from target-sequence complexes; (3) dissociation of the target-sequence complex and, finally; (4) amplification of dissociated sequences via polymerase chain reaction (PCR) to prepare an enriched candidate sequence pool for the next screening cycle.[45, 61, 278] Minor variations ranging from the oligonucleotide library base length and chemical composition (e.g. RNA vs. DNA for the same target[279]) to the number and stringency of screening rounds are routinely reported; however, the inclusion of intermittent PCR-based library enrichment steps is

common to all SELEX-based aptamer screening approaches. While SELEX certainly enabled the revolutionary discovery of aptamers, this evolutionary selection approach presents key bottlenecks to the aptamer screening process. As one example, the inclusion of intermittent partitioning, dissociated sequence recovery and amplification steps during each round renders the SELEX process itself as time-consuming (~2-3 months) and laborious.[46, 278] To address this specific issue without compromising on the number of selection rounds, robotic systems have been designed by one of the pioneers of SELEX to automate handling steps.[280, 281] Another key drawback centers on the likelihood of biased amplification as well as the formation of undesirable side products during PCR of a heterogeneous sequence population unless PCR parameters (e.g. cycle number[66]) at each selection round (involving varying concentrations of dissociated sequences recovered from target-sequence complexes) is optimized. To reduce the propensity for introducing PCR-based artifacts into the screening library, recent studies report purifying PCR products each SELEX round[282]; incorporating modifications to SELEX such as mixing PCR-based enrichment selection rounds with PCR-free selection rounds[283, 284]; or simply reducing the selection process to a single target-library incubation round with a small candidate sequence library volume and variation in wash steps to promote recovery of higher affinity candidates[285, 286]. Additional modifications to the SELEX process to alleviate PCR bias and side product formation have included the implementation of emulsion PCR (ePCR)[69], isothermal amplification, [287, 288] assymetric PCR, [289] lambda exonuclease digestion,[290] biotin streptavidin separation,[291] size separation of PCR product with denaturing-urea PAGE, [292] or digital droplet PCR (ddPCR)[70].

In contrast to these above studies incorporating various modifications to SELEX, Tapp *et al.* [293] recently reported an alternative to SELEX screening called CompELS (Competition-Enhanced Ligand Selection). In brief, CompELS involves the repeated introduction of unenriched random sequence populations to a particular target during each cycle, but notably avoids all the intermittent dissociation of target-oligonucleotide complexes and PCR-based library enrichment steps inherent to SELEX. Thus, PCR is carried out a total of only two times as part of CompELS screening against the chosen target - one time to prepare the master single-stranded DNA screening library and then a second and final time following the completion of screening in order to insert sequences into plasmids for sequence identification. In addition to saving time, resources, and labor as a one-day screening process, this nonevolutionary aptamer screening approach sidesteps the likelihood of introducing PCR artifacts while intentionally promoting competition between target-bound sequences (remaining from a prior CompELS screening cycle) and soluble candidate sequences (introduced during a subsequent screening round).

While planar {111} gold was chosen as the target for studies reported by Tapp *et al.*[294], gold nanorods prepared using the seed-mediated approach of Nikoobakht and El-Sayed[199] serve as the target in the current studies. While most agree that for AuNR synthesized with this approach contain {100} and {110} facets, quantitative assignments of crystallographic facets for gold nanorods is still under intense debate [184, 202, 295-297]. These displayed facets can serve as potential DNA aptamer binding sites along the rod-shaped gold nanoparticles. In addition to providing continuity in the chemical composition choice of our prior planar and current colloidal targets for CompELS-based aptamer screening and analysis, gold nanorods represent a well-studied colloidal system

with unique spectroscopic signatures depending on the size, aspect ratio, and dispersion state of gold nanorods in a suspension.[184, 198, 275, 298]

3.2 Material and Methods

3.2.1 Materials

Random libraries consist of single-stranded DNA (ssDNA) templates 69 bases in length with a central 40 base random segment region between fixed base segments on the 5' and 3' ends to enable primer binding for PCR (5'-GGG ACA GGG CTA GC-[40N]-GAG GCA AAG CTT CCG -3'). Normal libraries composed of an approximate equivalent distribution base (25% A, 25% C, 25% T, 25% G) and A-rich libraries (40% A, 20% C, 20% T, 20% G) were purchased from Integrated DNA Technologies (IDT, Coralville, IA). A-rich sequences were included based on prior studies indicating adenine-rich oligonucleotides have stronger interactions with gold surfaces.[122] Random libraries were PAGE purified by the manufacturer while the phosphorylated reverse primer (5'-Phos-CGG AAG CTT TGC CTC-3') and forward primer (5'-GGG ACA GGG CTA GC-3') ordered from IDT were HPLC purified.

dNTP mix (10 mM), phenol:chloroform:isoamyl alcohol (25:24:1), ethidium bromide, TOPO TA Cloning Kit for Subcloning, One Shot TOP10 Chemically Competent *E. coli* and X-gal were purchased from Invitrogen (Grand Island, NY). GoTaq DNA polymerase and 5X colorless GoTaq reaction buffer were purchased from Promega (Madison, WI). $CaCl_2$, HEPES, $MgCl_2$, $CaCl_2$ and KCl were purchased from BDH Chemicals (VWR

Scientific, Radnor, PA). SOC medium, Ethanol and Tris EDTA pH 7.4 were purchased from Fisher Scientific (Pittsburgh, PA). Hexadecyltrimethylammonium bromide, gold (III) chloride hydrate, silver nitrate, L-ascorbic acid, ampicillin sodium salt, and dimethylformamide were purchased from Sigma Aldrich (St. Louis, MO). Sodium borohydride was purchased from Fluka Analytical (Munich, De). Agar (bacteriological), glycogen and LB broth (Luria-Bertani) were purchased from Amresco (Solon, OH). Lambda exonuclease enzyme and 10X lambda exonuclease reaction buffer were purchased from New England Biolabs (Ipswich, MA). The MinElute PCR Purification Kit was purchased from Qiagen (Gaithersburg, MD). The pH 7.4 aptamer binding buffer (ABB) used for CompELS-based selection consists of 20 mM HEPES, 2 mM $MgCl_2$, 150 mM NaCl, 2 mM $CaCl_2$ and 2 mM KCl. All buffers and synthesis were prepared using 0.2 μm filtered water (18 MΩ-cm) from a Barnstead Nanopure ultrapure water purification system (Barnstead, Thermo Fisher Scientific, Inc., USA).

3.2.2 Synthesis of Gold Nanorods

A seeded gold nanorod (AuNR) synthesis approach was undertaken to prepare target suspensions.[199, 299] A CTAB solution (5 mL, 0.2 M) and a $HAuCl_4$ solution (5 mL and 0.5 mM) were mixed with a magnetic stirbar for 30 minutes at room temperature to feed a seed solution. At the 30 min time point, a fresh solution of 10 mM $NaBH_4$ was prepared in ice cold water, and 600 μL was added to the seed solution. The magenetic stir speed was increased to high for 2 min, allowing the solution to change to brownish yellow. For the nanorod growth solution, a CTAB solution (20 mL, 0.2 M) was mixed with a freshly prepared $AgNO_3$ solution (640 μL, 5 mM,) and then a $HAuCl_4$ solution (20 mL, 1 mM) under gentle magnetic stirbar mixing conditions for 2 minutes. Ascorbic acid (216 μL, 0.1

M) was then added to the nanorod growth solution and continuously stirred until the solution became colorless. Seed solution (48 μL) was then added to the nanorod growth solution and mixed with a magnetic stirbar for 40 min. To inhibit further growth or aging of the AuNRs and to remove excess CTAB from solution, the nanorod solution was centrifuged at 21.1 kG for 30 min and followed by removal and AuNR resuspension and they were resuspended in 40 mL of nanopure water; to complete one wash step. This wash step was repeated once more for a total of two wash steps. 2X washed nanorods were aged for 3 days at room temperature in preparation for CompELS aptamer screening.

3.2.3 ssDNA Library Preparation

Template sequences were amplified via polymerase chain reactions with either the random or A-rich template (0.17 pM), dNTPs (0.2 mM), forward primer (60 nM), reverse primers (60 nM), GoTaq polymerase (0.05 U/μL), and 1X supplied colorless GoTaq buffer. PCR was carried out on a G-Storm thermocycler with a 100 °C heated lid, with a 2 min hold at 95 °C followed by 25 PCR cycles (30 s denaturation at 95 °C; 30 s annealing at 47 °C; 30 s extension at 72 °C), and a final hold at 4 °C. An ethanol precipitation was performed on the resultant PCR product. Resuspended PCR product was digested with lambda exonuclease at 5 U/μg following the manufacturer's instructions to remove the phosphorylated hybridization partners. P/C/I extraction was performed on the digested PCR product and followed with another ethanol precipitation. Final ssDNA product was resuspended in aptamer binding buffer (ABB) and ssDNA concentration was adjusted to 2.5 μM and stored at 4 °C until used for screening.

Table 3.2.1 List of template DNA sequences with a central 40 base-long randomized segment (40N) flanked by fixed base segments (underlined) as well as forward primers and phosphorylated reverse primers used to prepare normal (i.e. equibase) and A-rich (i.e. 40% A; 20% each of C,G,T) random sequence libraries via solution PCR followed by lambda exonuclease digestion of phosphorylated hybridization partners to prepare the master, single-stranded DNA screening library. For cloning, non-phosphorylated reverse primer was employed.

DNA	Sequence
Template	5'-GGGACAGGGCTAGC-40N-GAGGCAAAGCTTCCG-3'
Reverse Primer	5'- CGGAAGCTTTGCCTC-3'
Forward Primer	5'-GGGACAGGGCTAGC-3'

3.2.4 Aptamer Screening with AuNR targets

ssDNA library was separated into 10 aliquots of 100 μL in PCR tubes and denatured in the thermocycler with heated lid (100 °C), 90 °C for 10 minutes, 4 °C for 15 min, and 24 °C for 5 min. 142 μL of 2X washed AuNRs were aliquoted into a PCR tube and centrifuged at 21.1 kG for 30 minutes and the supernatant was removed (wash step). AuNRs were resuspended in 100 μL of 2 μM dNTPs in ABB and incubated for 30 min on rotomixer. After the 30 min incubation with dNTPs, another wash step was performed. At this point the AuNRs had irreversibly aggregated into a visible pellet. A single aliquot of ssDNA library was added to the AuNRs, incubated for 30 min on rotomixer, and followed by a wash step. This was repeated until all 10 aliquots had been incubated. Following the last of 10 target-library incubation and wash, 200 μL of ABB was added to the AuNR suspension followed by centrifugation and supernatant removal. These wash steps with ABB were repeated for a total of six washes. Following the six washes, nanopure water (74.7 μL), 5X GoTaq buffer (20 μL), dNTPs (2 μL, 10 mM), non-phosphoylated reverse primer (1.2 μL, 5 μM), forwared primer (1.2 μL, 5 μM), and GoTaq (1 μL, 5U/μL) were

added to the AuNRs. PCR cycling was carried out as detailed previously for the ssDNA library preparation. After cycling, dsDNA PCR product was stored at 4 °C until purification for vector insertion.

3.2.5 Vector Insertion, Sequencing, and Cloning

A Quiagen (Gaithersburg, MD) PCR purification kit was used and purified product was resuspended in 10 uL of 10 mM Tris-HCl pH 8.5. The 10 uL of purified PCR product was placed on ice and salt solution (2.5 uL, 1.2M NaCl, 10 mM MgCl2) and 2.5 uL TOPO vector from TOPO TA Cloning Kit (Invitrogen, Life Technologies, Grand Island, NY) was added. This ligation mixture was placed on a thermomixer at room temperature for 15 min at 500 rpm. 5uL of the ligation mixture was added to the TOP cells, gently mixed, and put on ice for 2 h. TOP cells and ligation mix are then heat shocked at 42°C for 30 seconds. 250 uL of SOC medium was added to each vial. Resulting cell suspensions were incubated on a shaker table incubator incubator for 1 hat 37 °C and 250 rpm. Transformed bacteria are then plated on LB-agar medium supplemented with ampicillin and X-gal, followed by overnight growth in a 37 °C incubator. Twenty-one positive bacterial colonies resulting from CompELS using each of two screening libraries (i.e. normal or A-rich random ssDNA libraries) are randomly picked and after standard plasmid purification following the manufacture's directions, were sent to GENEWIZ Inc. for sequencing analysis (South Plainfield, NJ).

3.2.6 Checking Recovered Sequences for Errors

Identified sequences returned from GENEWIZ Inc. were checked for errors in their fixed base segments, random region length (e.g. missing a base in the 40 base-long segment), or

non-determinate bases. Any sequence possessing one of these errors was removed from consideration. Any non-erroneous complementary sequences to the template provided as a recovered sequence reverse transcribed (analytically) to yield the original template.

3.2.7 Primary Structure Analysis

The primary sequences of identified aptamer candidates are analyzed and aligned using T-Coffee multiple sequence alignment of the 40 base central randomized regions (bases 15-54) of the sequences to identify and compare position-dependent bases among all aptamers selected from both the normal and A-rich libraries. Alignment results were inserted into Microsoft Excel 2016 and analyzed and color-coded using a macro. MEME 4.9.1 (freely accessed in 7/20/2017 at http://ebi.edu.au/ftp/software/MEME/) was then used to identify position-independent sequence segment motifs that occur anywhere within the central 40 base variable segment of the aptamer sequences.[172, 175]

3.3 CompELS Screening Results and Discussion

3.3.1 Preparation and Growth of Gold Nanorods

Gold nanorods (AuNRs) were analyzed via UV-Vis spectroscopy on the third day post synthesis and 2X washing as shown in Figure 3.3.1. Based on spectra peak values of 513 nm and 674 nm, AuNR was estimated to have an aspect ratio of ~2.7 [185]

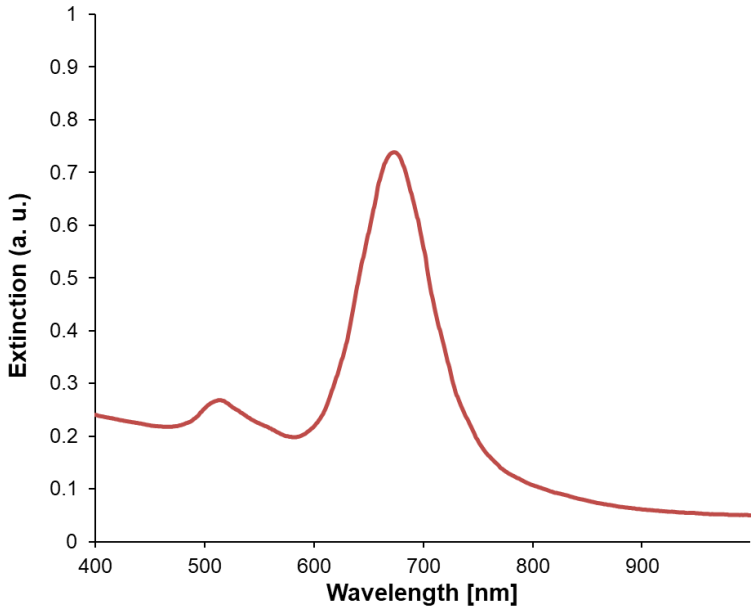


Figure 3.3.1. UV-Vis spectra of AuNR used for CompELS screening. For spectroscopy studies, AuNR were prepared using a seeded growth, washed two times to remove excess CTAB while maintaining a relatively stable suspension exhibiting peaks values of 513 nm and 674 nm. Prior to the start of CompELS screening, AuNR suspensions were washed an additional (third) time on day 3.

3.3.2 Competition-Enhanced Ligand Selection (CompELS)

Aptamer selection was performed in three separate CompELS screening sessions against the gold nanorods (AuNR) using equidistribution of bases ssDNA random libraries for the first two screenings (sequence sets 1XX and 2XX) and using an A-rich random ssDNA based library for the last screening (sequence set 4XX). As mentioned previously, the motivation for using A-rich screening libraries stems from prior work: studies not shown in chapter 2 of the aged gold seed solutions aggregating in with pure polyadenine. [122] The key differences in Competition Enhanced Ligand Selection (CompELS) from Systematic Evolution of Ligands by Exponential Enrichment (SELEX), as stated previously, is the removal of the intermittent dissociation of target-oligonucleotide

complexes and PCR-based library enrichment steps. To minimize potential non-specific binding of ssDNA to the AuNR, AuNR were first incubated with dNTPs at 2 µM. After 30 min of incubation with dNTPs, the supernatant from the nanorods was removed and their irreversible aggregation and formation of a small pellet in the bottom of the PCR tube was observed. Addition of a prepared 2.5 µM ssDNA random library was introduced to the target to begin CompELS as shown in Figure 3.3.2.

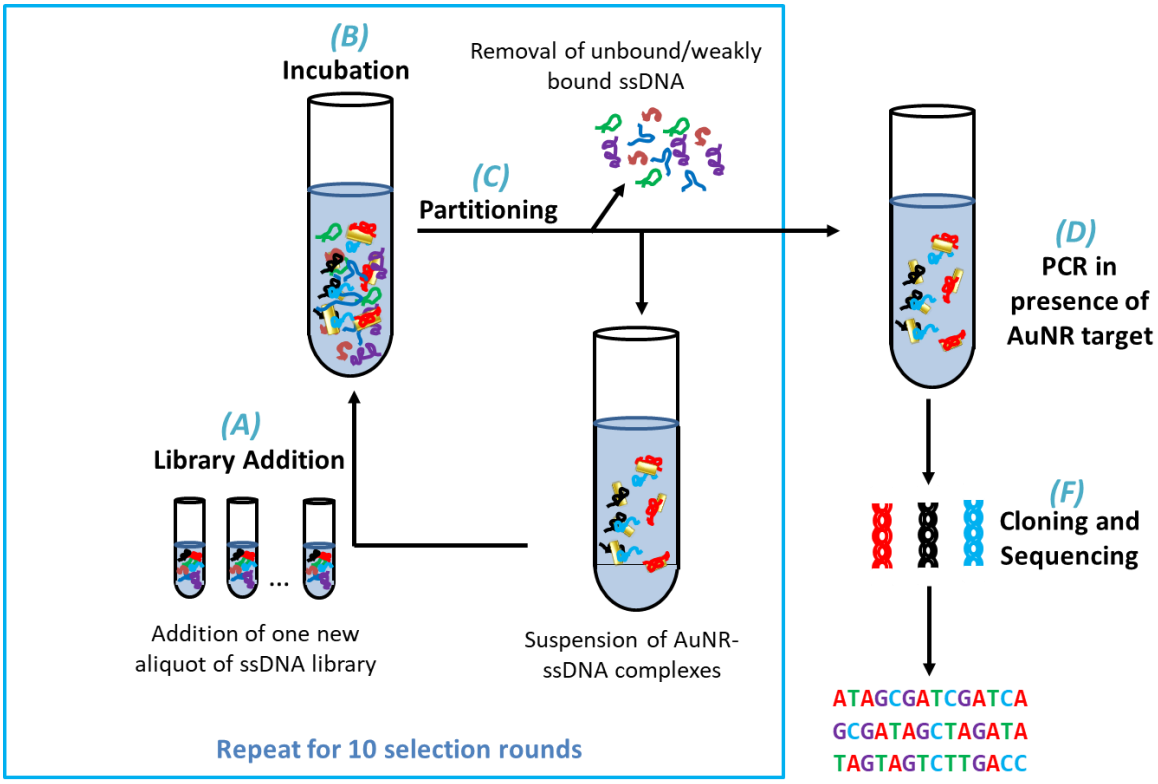


Figure 3.3.2. Schematic of CompELS aptamer selection for DNA aptamers against gold nanorods (AuNR) that entails (A) addition of one ssDNA library aliquot to (B) AuNR followed by (C) removal of unbound, weakly bound, or displaced DNA sequences. Following the completion of 10 rounds of CompELS, (D) PCR is carried out in the presence of AuNR and finally (E) cloning and sequencing of aptamer candidates is undertaken.

The concentration ratio of ssDNA to particles chosen was to ensure that ssDNA was well in excess of the surface available to the particles with the number of particles being used

in each panning to be ~10^{10}. Previous characterization by TEM of synthesis (not shown) showed a ~21 nm diameter of formed rods with our seeded synthesis, estimated surface area available to be ~5000 nm^2 per gold nanorod in solution. Estimating the size of a random coil ssDNA sequence from the library using a freely jointed chain model for a sixty-nine base sequence yields a radius of gyration, R_g, of ~3.32 nm. Based on these assumptions, the number of ssDNA sequences used for panning in each CompELS round was ~100X in excess of that required to form a monolayer adsorbed, 2D close-packed of random coils of ssDNA on all gold nanorods in solution. After each incubation, non-binding DNA sequences in the supernatant were removed, and a new aliquot of ssDNA library was introduced to the AuNR for the next CompELS round in the presence of previous adsorbates. Notably, this step differs from SELEX in which bound sequences are eluted, recovered, and then amplified via PCR during each round. After 10 selection rounds, six washes of the AuNR with the binding buffer were used to remove any remaining weakly associated adsorbates. Lastly, PCR is performed in the presence of the gold nanorods to amplify bound sequences and allow for recovery, vector insertion, cloning and sequencing. A total of 42 aptamer sequences were identified from the three CompELS screenings with the number of sequences being identified in each screening being 15, 8, and 19. Recovered aptamer sequences are shown for the normal random library in Table 3.3.1 and for the aptamer sequences from the A-Rich library in Table 3.3.2.

Table 3.3.1. AuNR aptamer candidate sequences from normal libraries discovered via CompELS screening

Aptamer Nomenclature	Sequence (5'→3')
101	GGGACAGGGCTAGCATATGTAATTATGCGCTTCTAGTTAATAGGCCCTTTGCAAGAGGCAAAGCTTCCG
102	GGGACAGGGCTAGCATTAATTTAAGCTTCATCAGACAAGAGCAGGGGCATACAAGAGGCAAAGCTTCCG
105	GGGACAGGGCTAGCATATGATCGTTTATAAGCTGTTCTTCTTCCATGGTTACATGAGGCAAAGCTTCCG
108	GGGACAGGGCTAGCAGTGTTACAGTGTTGGGTTTTTTAATAACAATGGATTTGTGAGGCAAAGCTTCCG
110	GGGACAGGGCTAGCGTTTAATACAGGTTCACGAGGCATTACGTTTACTTACTCTGAGGCAAAGCTTCCG
111	GGGACAGGGCTAGCTGCAGCGACGGGTTATTATTCTTGTACAGTTCTTTTTCACGAGGCAAAGCTTCCG
112	GGGACAGGGCTAGCTAATACATAACGTAGTTATTCCGCATTTAACTAGAAAATTGAGGCAAAGCTTCCG
114	GGGACAGGGCTAGCATGATAACCTATTTTGCATTTATATGGTGCAGAACCATTTGAGGCAAAGCTTCCG
115	GGGACAGGGCTAGCAACGAAGTCTGGGTAGAGGCGTGATGTGGGATGTCTAGGTGAGGCAAAGCTTCCG
116	GGGACAGGGCTAGCATCTTTCTCGTTTGAAAACCTCGGCTCTTTGCCATGGGCAGAGGCAAAGCTTCCG
117	GGGACAGGGCTAGCTAGGGCGTGGGTGGTGTTTAATATTCCCTATGCCCGACGTGAGGCAAAGCTTCCG
118	GGGACAGGGCTAGCTATTTGTTGGTCTTTAAAAAGTGTTACACTCGCTATTGCTGAGGCAAAGCTTCCG
119	GGGACAGGGCTAGCACGTAGAGGGGGCTTGTCGTTAACAAAATGGGGTTCTGGCGAGGCAAAGCTTCCG
120	GGGACAGGGCTAGCGAAATTAGATCATGCGGCTAAGGGGTTAGTTAAGTATCCAGAGGCAAAGCTTCCG
121	GGGACAGGGCTAGCAAAAATTATGATTTGTGAGTGAACTCATTCCCATTGACATGAGGCAAAGCTTCCG
201	GGGACAGGGCTAGCCGTAAATGTCTCTGAACCTTAAACTGTCAGGCAGATTTATGAGGCAAAGCTTCCG
204	GGGACAGGGCTAGCACGTCGTTGTATGGGACTATTGTAGTAAACTTCAATTATTGAGGCAAAGCTTCCG
207	GGGACAGGGCTAGCTAGTTATGGAACGCAATTGGGGGACGGATCTCTTGTCTTTGAGGCAAAGCTTCCG
210	GGGACAGGGCTAGCTGCCATTCCGAAACGAAATAAAAGTAACCCTTATTTTCACGAGGCAAAGCTTCCG
212	GGGACAGGGCTAGCCGGGGCTGTTGGGGGCAAATCATTTAGTCATTGAGGTATCGAGGCAAAGCTTCCG
214	GGGACAGGGCTAGCGTTATATTTTAAAAGTGTCCATTTGGGCGAATTATGGATAGAGGCAAAGCTTCCG
217	GGGACAGGGCTAGCACCTGATGTTGATGTGTATCCTCCGTCGACAGTATATGGTGAGGCAAAGCTTCCG
219	GGGACAGGGCTAGCATTAGTTTGGTTGTATGTGTTACAGGATTGTAGGGGATGTGAGGCAAAGCTTCCG

Table 3.3.2. AuNR aptamer candidates from A-rich libraries discovered via CompELS screening

Aptamer Nomenclature	Sequence (5'→3')
401	GGGACAGGGCTAGCATATGGATAAGATCTGCAGGCAGAATCCAGTAGTTAATTAGAGGCAAAGCTTCCG
402	GGGACAGGGCTAGCACACGTAAAAAGGTGTGAATCGGTGATGAAGAGGTTTTTCGAGGCAAAGCTTCCG
403	GGGACAGGGCTAGCAAACGAAGAGATTATAGATGAGATACCGTCTCGAAACCACGAGGCAAAGCTTCCG
404	GGGACAGGGCTAGCGGCAAATAACTGGACGTACTAAGAACAATAAGCCTGGGTAGAGGCAAAGCTTCCG
405	GGGACAGGGCTAGCGGATACGTTAACCAGTACTAAAAGCGAGCATACCTAAAAAGAGGCAAAGCTTCCG
406	GGGACAGGGCTAGCGCAAAGAAACTAGGATAAAAGGAGCCAGGATTCACAAATAGAGGCAAAGCTTCCG
409	GGGACAGGGCTAGCAAATCGTGAATTTATAGGTATAAACGTGAAATACACAAATGAGGCAAAGCTTCCG
410	GGGACAGGGCTAGCTCTGACAACTTCACTACACGCAACGGGAAAAGATATTGGAGAGGCAAAGCTTCCG
411	GGGACAGGGCTAGCCGTTATGAGAAGCAACTTCCTGGCAATAACTAGGAGCCAAGAGGCAAAGCTTCCG
412	GGGACAGGGCTAGCAGGTAAGTTATTTGAGATGTCATGTAAGGACAAAGTGGTAGAGGCAAAGCTTCCG
413	GGGACAGGGCTAGCTTAACCCGATCACAGTCCGACTCATTAACATATTAGTGTAGAGGCAAAGCTTCCG
414	GGGACAGGGCTAGCATGGATTGCCGAGAACTCCACAAGATCTTAATCGAATGACGAGGCAAAGCTTCCG
415	GGGACAGGGCTAGCAGATAATCAGTCACTAGAAAAAATCCCAGCACCAGGAGTAGAGGCAAAGCTTCCG
416	GGGACAGGGCTAGCGATATGACGAAGCAAATGTCATTAATAACAAACAAAGATCGAGGCAAAGCTTCCG
417	GGGACAGGGCTAGCAAGGGTAAAATGAATTTAACATAGTATTAGAACGTGTCACGAGGCAAAGCTTCCG
418	GGGACAGGGCTAGCAAGCCTTAAAGGAAACGGACGAAAAAATAAATTAGGTTATGAGGCAAAGCTTCCG
419	GGGACAGGGCTAGCTTTTAGCCATAGAGTACATACGACTCCGGAAAGAATTCAAGAGGCAAAGCTTCCG
420	GGGACAGGGCTAGCTGATAATTAAATTATACCCCAGCTATTGTTTACATCGCACGAGGCAAAGCTTCCG
421	GGGACAGGGCTAGCAAAAGAATAAACGACTGAAAAATAAACGGTTAAGAAGTGTGAGGCAAAGCTTCCG

3.3.3 Alignment of Gold Nanorod Aptamer Sequences

The identified aptamer sequences were analysed to identify base position dependent similarities in their primary structure as well as searching for base position independent similarities in primary structure or base sequence. Figure 3.3.3 shows the central random region (bases 15-54) of all identified aptamers. Similar to a heat map, highest percentages of base identity matches are indicated by red while lowest percentages are indicated by blue. Of these base positions, stronger consensus (greater than 50%) is only base 15. Given that some of these sequences were A-rich and the base with above 50% consensus is an

adenine, this was determined that this was not statistically significant; this was later confirmed by T-Coffee alignment.

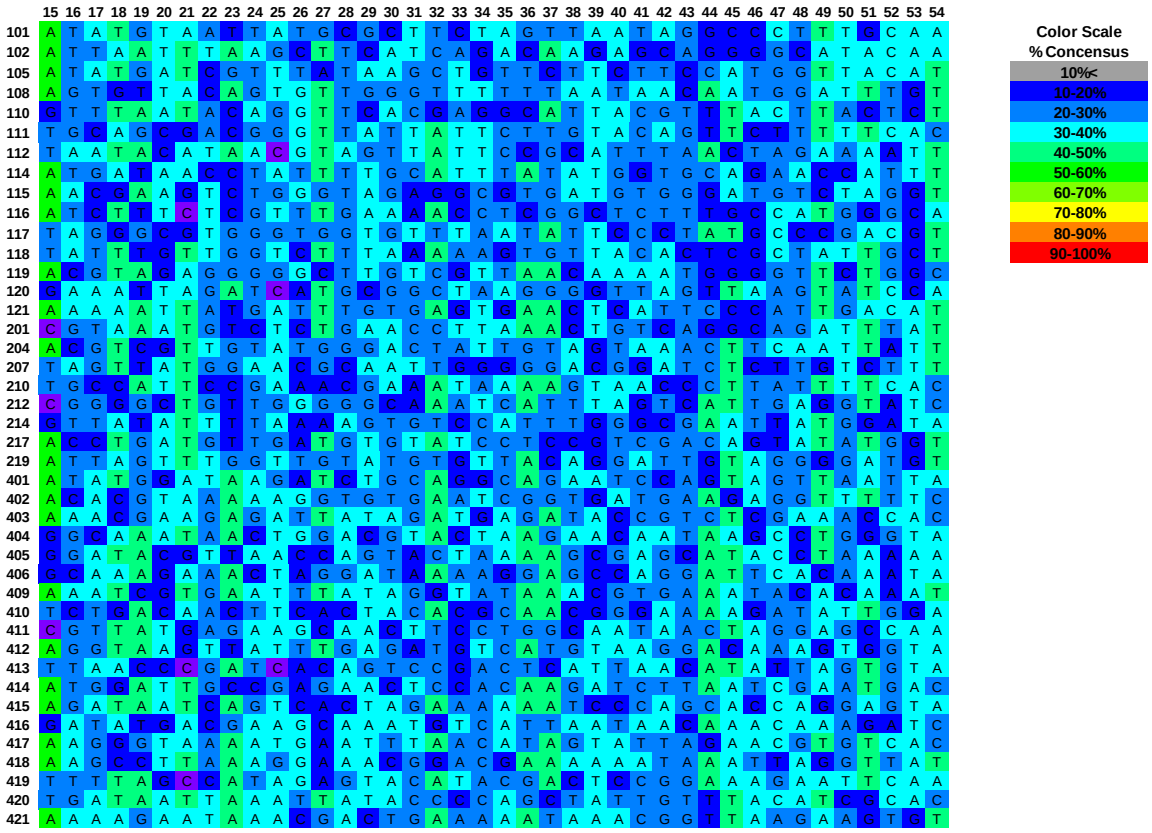


Figure 3.3.3 Conservation of base identity at all central 40 base positions (15-54) for all aptamer sequences. Sequence nomenclatures are shown in the left column with their associated sequence being shown in that row. Base 15 of this central region is the only base position that has a conservation greater than 50% in the random region with 22 sequences sharing that adenine base at that position. Identity overlap which adenine as the shared base identity across all 42 aptamer sequences.

In order to determine if other primary structure similarities in terms of base positions were present, an alignment was performed using T-Coffee as shown in Figure 3.3.4 to allow gaps to be inserted before, after, or between bases to potentially allow more primary structure comparisons to be made. Even with the alignment by T-Coffee only two newly indexed base positions contain above 50% conservation, as compared to one with no

alignment. The T-Coffee alignment showed no significant relationship between the sequences (analysis portion not shown).

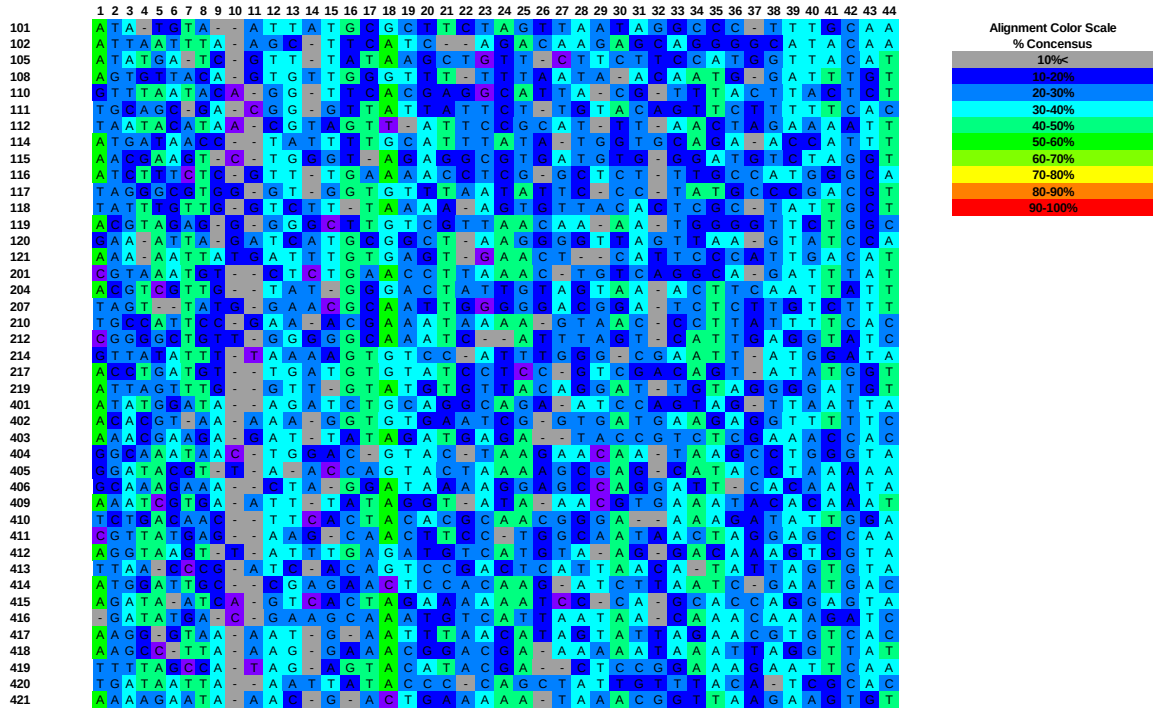


Figure 3.3.4. T-Coffee alignment of all aptamer sequences color-scaled to show consensus percentage. All gap insertions are indicated with a dash (total of 4). Preservation of an adenine base greater than 50% is only observed at the new index locations of 1 and 18.

3.3.4 MEME Analysis of Gold Nanorod Aptamer Sequences

Multiple EM for Motif Elicitation (MEME) was performed on the sequences to identify base segment motifs that can occur at different base positions between two or more aptamer sequences. This analysis determines if significant structural repeats and assigns it an expectation value or E-Value that relates the probability of observing a particular motif without gaps in given sequences over the probability of observing the motif randomly. An E-value greater than 0.01 is not normally considered statistically significant.

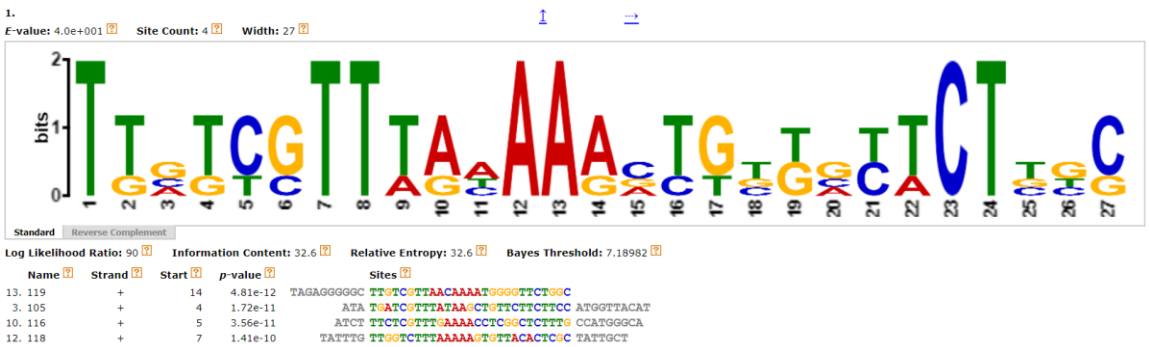


Figure 3.3.5. Motif logo from MEME analysis of all AuNR aptamers. Lowest E-value yet not statistically significant value of 4.0e+001. At the top of this figure, MEME analysis shows the E-value of, the number of sequences or sites the motif was found (4), and the width of the motif (27 bases). Sequences involved in this non-significant motif were 105, 116, 118, and 119.

Based on the fact that the MEME analysis does not return E-values of statistical significance (with all being greater than or equal to 4), no consensus motifs appear present among the CompELS identified aptamers. This finding is not entirely surprising since CompELS, unlike SELEX is not an evolutionary screening process.

3.4 Conclusions

While identification of aptamer sequences with the CompELS process was successful, apart from an adenine overlap in the first base identity for more than 50% of aptamers, no patterns in primary structure emerge with the current analytical tools available in either T-coffee alignment of all aptamer sequences or with MEME based motif searching among subsets of aptamer sequences. Since CompELS is inherently a nonevolutionary approach that avoids intermittent PCR steps, there are fewer opportunities to, for example, introduce mutations or otherwise evolve similar primary structures among sequences over several "PCR generations." In order to assess similarity or motifs of aptamers generated through

nonevolutionary routes, other methods of analysis that focus on secondary structure analysis are addressed in the next chapter.

CHAPTER 4. EXPANDING ANALYITICAL TOOLS FOR DNA SECONDARY STRUCTURE ANALYSIS OF APTAMER CANDIDATES IDENTIFIED THROUGH A NONEVOLUTIONARY SCREENING APPROACH

In this chapter, computational methods for protein sequence (i.e. primary structure) analysis have been adapted and developed to identify potential patterns or shared "motifs" in secondary structure compared to a random sequence population employed as a background "signal" for normalization purposes. Methods developed include initial manual analysis that transformed into a more high-throughput secondary structure family (SSF) identification as well as base pair mapping across all base positions, base-position dependent SSE quantification followed by complimentary position specific score matrixes (PSSM) to provide a more weighted SSE quantification approach, and finally multi-secondary structure alignment (MSSA) to strategically insert gaps and thus establish the distribution of particular domains possessing a string of predominate secondary structure elements. Cumulatively from these analyses, some distinctive secondary structure patterns emerge to distinguish aptamers from a random sequence population.

4.1 Introduction

Explosive progress in sequencing tools such as Next Generation Sequencing has been met by advances in sequence analysis tools to identify consensus motifs among subgroups of

sequences[300, 301] as well as to identify evolutionary connections between larger groups of sequences using global alignment[170, 302, 303] approaches. Among the multiple sequence alignment tools, ClustalW[170] endures as one of the most cited[304] while updated algorithms continue to refine the alignment process[305, 306] to accommodate increasingly larger genomically or biologically-relevant sequence data sets as well as groups of aptamer sequences selected via SELEX[45, 307]. In contrast, aside from schematically illustrating predicted self-hybridized structures of aptamer sequences using online tools such as UNAFOLD[18], the importance of relating secondary structures of DNA aptamer sequences to specific target binding is commonly stated, but generally less examined in the literature. The exceptions to this trend tend to reside with historically popular DNA aptamers such as Bock's aptamer for thrombin[308-310] or with studies focused on identifying potentially shared binding motifs[47, 311] among different DNA aptamer sequences for the same target. Such structural information can allow, for example, for guided aptamer sequence truncation efforts[311-313] or strategic extension of hybridized segments to stabilize self-folded structures[311] rather than arduous trial-and-error experiments to unmask the binding motifs in shortened, often higher affinity[120, 314] aptamer segments. Moreover, comprehensive three-dimensional structural databases for RNA alone[315, 316] and RNA-protein complexes[317, 318] continue to expand to inform structural analysis and design strategies for RNA-based aptamers. Knowledge of structure-binding relationships in DNA aptamers, on the other hand, lags in comparison to RNA aptamers. This lag can be attributed, at least in part, to two inter-related challenges: (1) the nontrivial nature of these characterization pursuits (as demonstrated with multidimensional NMR spectroscopy studies of RNA aptamers[319, 320]) and (2) the lack

of three-dimensional databases and predictive tools with only a recent exception for self-hybridized DNA sequences to help inform structural characterization studies of DNA aptamers alone and bound to their specific target.

To begin closing this analytical gap in understanding the specific role that structure may play in DNA aptamer-target binding, Tapp [293] defined a new structural classification scheme to identify shared features and patterns in predicted minimum or lowest free energy secondary structures among DNA aptamer sequences identified from CompELS screening. Here, the current work expands on secondary structure analysis to include relevant suboptimal secondary structures of all aptamer candidates identified from CompELS and adapts an existing multiple protein sequence alignment tool in order to identify shared domains within aptamers where particular secondary structure features predominate. Notably, while analytical tools for aligning pairs[321, 322] as well as groups[323-325] of RNA secondary structures and even suboptimal secondary structures[326] are available, these tools generally presume an evolutionary connection between biologically-relevant RNA sequences[327] and, depending on the alignment program, take into account secondary structure features such as pseudoknots[328] specific to RNA sequences. These algorithms have now been extended to identify secondary structural motifs in RNA aptamers identified via conventional SELEX[59, 329] as well as high throughput[181, 330] SELEX. In light of the nonevolutionary nature of the CompELS screening process used in the current work as well as the structural features that are unique and better catalogued in previously mentioned RNA databases[315-318], the analytical approach reported here for DNA aptamers for AuNR strives to avoid introducing potential artifacts while still leveraging the capabilities of existing, well-cited analytical tools, position specific score

matrixes (PSSM)[171, 176] and ClustalW[170], to determine if any correlations or patterns in predicted secondary structures exist between nonevolutionarily-selected aptamers.

4.2 Materials and Methods

4.2.1 Sequences Used

4.2.1.1 Aptamer Sequences

DNA aptamer sequences identified via CompELS against AuNR targets as detailed in and listed previously in Chapter 3 in Table 3.3.1. AuNR aptamer candidate sequences from normal libraries discovered via CompELS screening, and in Table 3.3.2. AuNR aptamer candidates from A-rich libraries discovered via CompELS screening, were used.

4.2.1.2 Generation of Model Random Library

A random population of 1000 DNA sequences, each with a randomized 40 bases long central segment flanked by the same fixed base segment in the AuNR aptamers was generated for use as a model background sequence for secondary structure analysis in Microsoft Excel 2016 using the built in random function =RAND() to create a random array of 40 by 1000 with random value between 0 and 1. Excel's random number generator has been shown to be adequate in other studies.[331, 332] The random region of 1000 sequences were simulated with the function:

=IF(#REF<0.25,"A",IF(#REF!<0.5,"C",IF(#REF!<0.75,"G","T"))) (4.2.1)

was used to generate the central randomized region of bases in which "#REF!" refers to a cell at *column#, row#* located in the previously generated random array. This generated

central random region was then inserted into the same template design (i.e. same 14 and 15-base long fixed base segments at the 5' and 3' ends) as the sequence library used for CompELS, as shown in Table 3.2.1. Initial analysis of the random sequence population indicates ~0.2% difference from an (ideally) equivalent (25%) distribution of A, T, C, G and base at each position in the 40 base long segment. Variance of individual base content across sequences was 7.4 .

4.2.2 Secondary Structure Generation

Zuker's UNAFOLD web server[15] was used to generate secondary structures of all sequences under conditions mimicking that of the aptamer binding buffer (ABB) of $[Na^+]$=152 mM, [Mg++]=4 mM at 23°C and to calculate the Gibbs free energy, ΔG, of self-hybridization. Secondary structures with the lowest predicted ΔG values were assigned S1 nomenclatures as the most thermodynamically favourable or *dominant* secondary structure. Other *suboptimal* secondary structures of the same sequence with a ΔG within a 5% difference of the dominant secondary structure ware also considered for analysis. These suboptimal secondary structures were assigned S2 through S8 nomenclatures structures in order of the lowest (i.e. most negative) to highest (i.e. least negative) predicted ΔG. A table with associated predicated ΔG values of self-hybridized structures of aptamers from UNAFOLD is available in the appendix. All dominant and suboptimal aptamer secondary structures, base pairs were recorded and transcribed by hand into Microsoft Excel for further analysis. Section 4.2.5 will discuss additional methods used to annotate secondary structures.

4.2.3 Base-Pair Mapping

Recorded base pair positions (base number X, base number Y) along each single-stranded 69 base long aptamer sequence were mapped. As base pairing through self-folding symmetrically mirrors about the start of a fold, the grid would, in essence, duplicate base-pairs across the "mirror" line drawn in each map as indicated in Figure 4.2.1. Rather than simply duplicating all the base-pairings across this mirror line in each grid, the bottom left half of the map was populated by all base pairings, only overlapping of "shared" base pairings between 2 or more self-hybridized sequences are shown in the top right half of the map. An example of this mapping format is shown in Figure 4.2.1, with 2 secondary structures from an aptamer sequence.

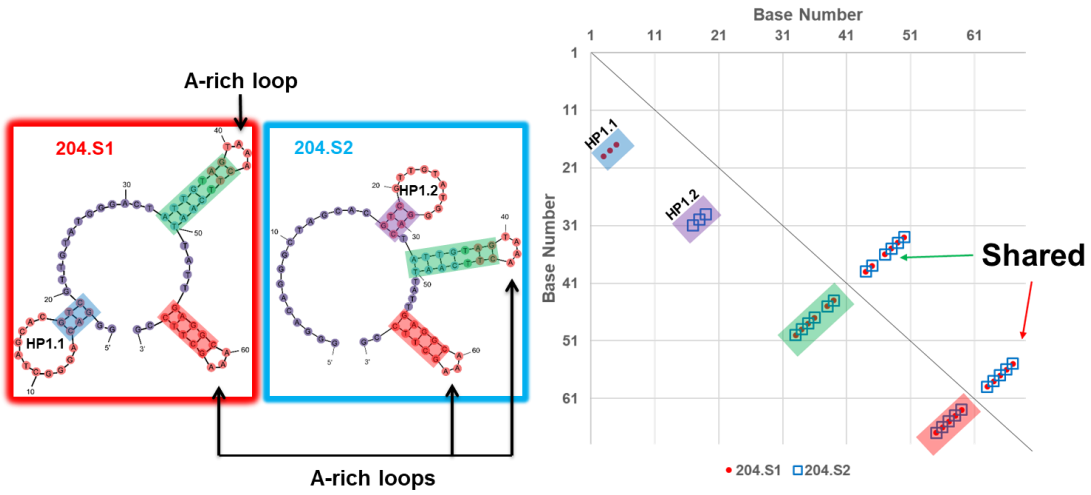


Figure 4.2.1 Two predicted secondary structures color-coded to indicate shared and unshared base-pairs indicated in the adjacent base-pair map for 204.S1 and 204.S2. These two secondary structures for the 204 aptamer sequence also belong to the same secondary structure family.

UNAFOLD performs similar base pair analysis for generated structures, but only performs comparison, as in the example shown in Figure 4.2.1, where only structures originating

from a single sequence are compared.[15] For our analysis this was extended to all considered structures and all sequences.

4.2.4 Secondary Structure Elements (SSE)

To enable grouping and classification of different predicted secondary structures, each predicted secondary structure for an aptamer sequence is broken down into 1 of 6 different types of secondary structure elements. These categories of secondary structure elements or SSE are shown in Figure 4.2.2 : single-stranded segment, hairpin, internal loop, bulge, duplex, and multibranched loop. A similar approach to structure classification was taken by Hoinka *et al.*[59] In later analysis, described in section 4.2.5, the hairpin was divided into two portions, namely the unhybridized loop and self-hybridized stem segment. This classification scheme also allows for conversion of the entire secondary structure for an aptamer sequence into text strings for analytical purposes as detailed next.

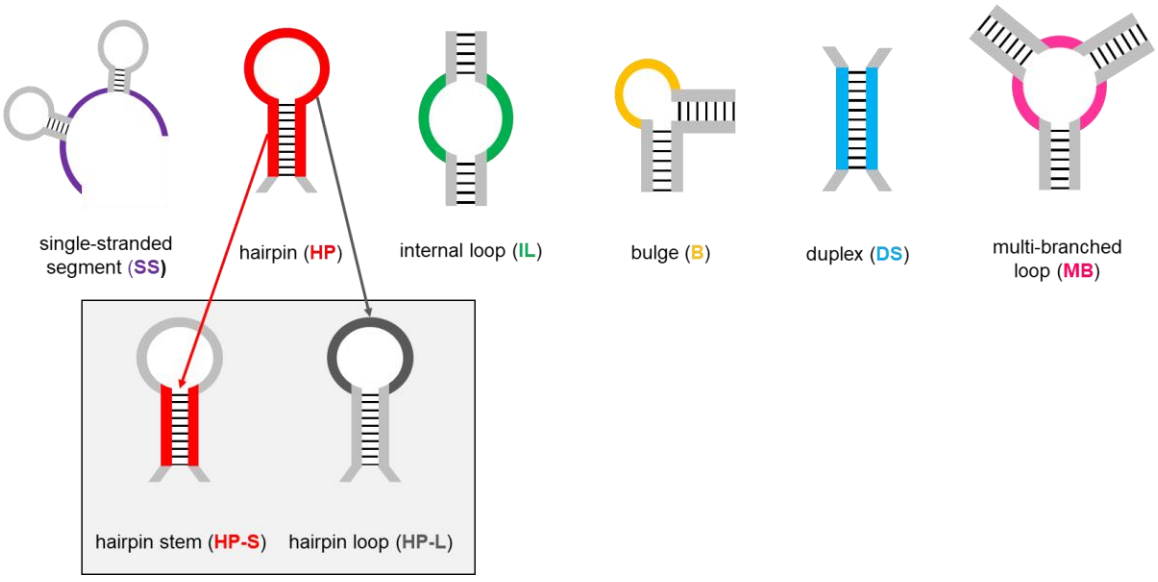


Figure 4.2.2. Schematics of six predicted secondary structure elements (SSE) used to classify secondary structure families (SSF). Also shown is the case where the hairpin was

divided into stem and loop portions for creation of secondary structure strings (SS$) used in base position structural analysis, position specific score matrixes (PSSM) and multiple secondary structure alignment (MSSA) analysis of AuNR aptamer sequences. Each element is color-coded as follows: purple for single-stranded segment, red for hairpin (red for just hairpin stem segment and dark grey for hairpin loop segment), green for internal loop segments, dark yellow for bulge segments, blue for double-stranded segments, and dark pink for multi-branch loop segments. Thin black lines represent Watson-crick or wobble base pair matches.

4.2.5 Conversion of Entire Secondary Structure from UNAFOLD into Secondary Structure Strings (SS$)

To first convert a secondary structure into a text string requires an alphabet. Each of the twenty natural amino acids has a one-letter designation and thus conveniently allows adaptation of existing protein analysis toolkits for the purposes of secondary structure analysis (involving 7 possible SSE). Table 4.2.1 lists how each SSE is assigned to an amino acid residue for converting secondary structures into text strings or secondary structure strings (SS$)

Table 4.2.1. List of amino acid identity assigned to each SSE for ClustalW analysis. For this analysis, a hairpin is broken down into its loop and stem components.

Abbreviation of Secondary Structure Element Notation		
Secondary Structure	**Shortened Notation**	**Amino Acid Substitution**
Single-Stranded Segment	SS	S
Hairpin Stem Segment	HP-S	H
Hairpin Loop Segment	HP-L	L
Internal Loop Segment	IL	I
Bulge Segment	B	G
Double-Stranded Segment	DS	D
Multibranched Loop Segment	MBL	M

An example of color-coding each SSE in a predicted secondary structure is shown in Figure 4.2.3 as follows:

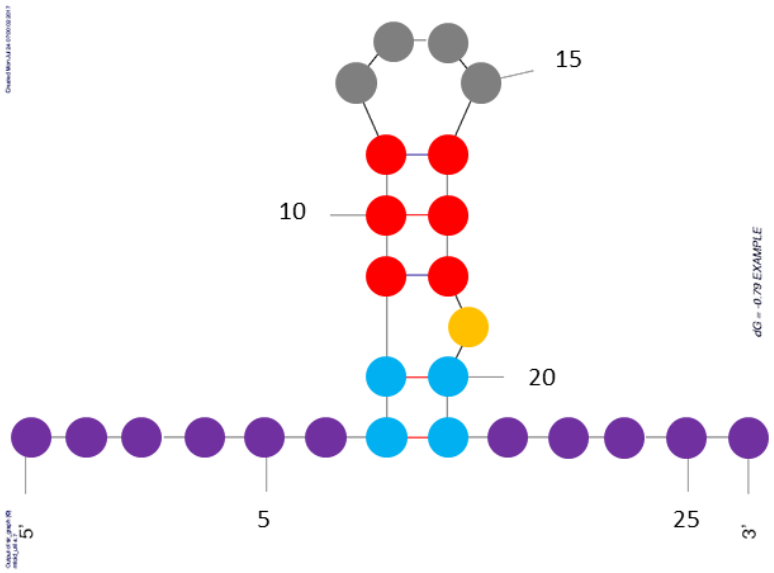


Figure 4.2.3 Example secondary structure showing single-stranded segments (purple), duplex (blue), hairpin stem (red), hairpin loop (grey), and bulge (yellow) elements.

To designate as a secondary structure string (SS$) starting at the 5' end, this same sequence is designated as follows:

SSSSSSDDHHHLLLLHHHGDDSSSSS

4.2.6 Secondary Structure Families (SSF) Analysis

A program in Matlab 2016 was created to determine the total number of the main 6 SSE associated with a particular sequence by counting the number of continuous segments of a given secondary structure element. For example, the above SS$ input of:

"SSSSSSDDHHHLLLLHHHGDDSSSSS"

would be reduced to the following:

SDHLHGDS

Using this reduced string, one can then count the total number of SSE in a sequence using the formulaic approach shown in Table 4.2.2

Table 4.2.2. Table of equations employed to count the total number of SSE in predicted secondary structure for 69 base-long sequences from SS$s. Revised equations were developed upon discovering Double-stranded segment-Bulge-Double-stranded segment (DGD) order of SS$s led to undercounting of double-stranded segments. Secondary formula shows that if N/2 + D/2 is not equal to an integer value, the computed value is rounded up to the next integer value to account for a counting error stemming from specific but unusual contiguous secondary structures (e.g. SDIHLHIDIHLHIDS).

Family Element Counting	Counting Basic Formula	Revised formula (DGD=N)
Number of Single-Stranded (#SS)	S	
Number of Hairpins (#HP)	L or H/2	
Number of Internal Loops (#IL)	I/2	
Number of Bulge (#B)	G	
Number of Duplex (#DS)	D/2	if((N/2 + D/2)=int., (N/2 + D/2), (N/2 + D/2 + 0.5)
Number of Multi-branched loops (#MBL)	0 or 1 step function	

As it can be seen in Table 4.2.2, the basic formulas for determining the number of secondary structure failed to correctly identify the number of SSE when more complex secondary structures were encountered. It was determined that the primary issue was with duplex-bulge-duplex regions. These regions would mask half of the presence of a duplex or double-stranded region as shown in

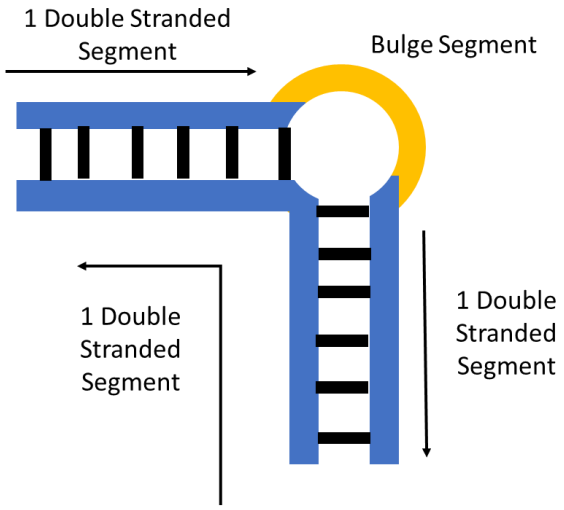


Figure 4.2.4 Masking of true count of number duplex from a SS$ by a central bulge.

A relationship was derived; it was determined that for every N bulge between duplex sections (DGD) that the number of duplexes associated with that bulge was N+1, and the number of duplex segments associated with the bulge was N+2. Thus, if DGD occurred, and D duplex segments not associated with the bulge were in the system, the number of total duplexes would be equal to:

$$Number\ Duplex(DS) = (N+1) + \frac{D-(N+2)}{2} = \frac{N}{2} + \frac{D}{2} \qquad (4.2.2)$$

The secondary equation in Table 4.2.2 stemmed from secondary structure similar to shown in Figure 4.2.5.

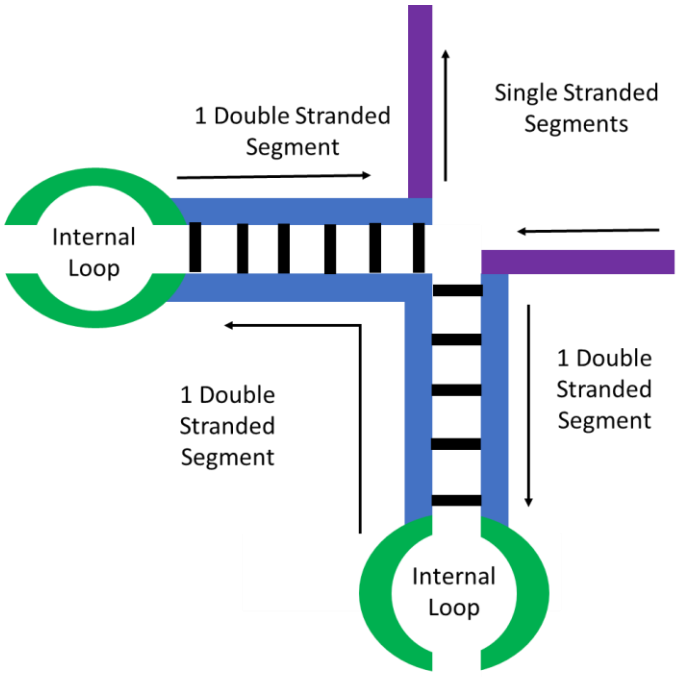


Figure 4.2.5. Example of specific SS$ of "SDI…IDI…IDS" with a continuous double-stranded segment over separate duplexes which masks the true number of duplexes, where "…" is another portion of secondary structure not shown next to the two internal loops to emphasize the important portions of the particular shown structure formation.

As our sequences were relatively short in base length for multiple occurrences of this secondary structure and this secondary structure was atypical, approximating with the revised equation shown in Table 4.2.2 is adequate, but all instances of non-integer values in the revised equation were checked manually.

Calculation of the number of multi-branch loops was more challenging. However, as the investigated sequences were short and the occurrence of more than one multi-branch loop is extremely improbable in these 69 base-long sequences, if any M segments were observed in the secondary structure string (SS$), those segments are considered part of one central

multi-branched loop occurring for the structure, and the number of MBL is considered to be one.

All sequences secondary structures are then classified into secondary structure families (SSF) based on how many of each SSE are present as indicated in the coordinate system (#SS, #HP, #IL, #B, #DS, #MBL). As an example, the secondary structure in, Figure 4.2.3 has the 6-cordinate value of (2, 1, 0, 1, 1, 0). While the 6-coordinate values are determined for each aptamer and random sequence, a SSF in the aptamer sequences must possess at least 2 members. Thus while every aptamer has a 6-coordinate set of values, not every sequence belongs to a SSF.

Following 6-coordinate value assignments for all aptamer and random sequences, one-tailed sided 2 proportion Z tests are undertaken, where applicable, to determine any statistically significant differences between the aptamer sequences and the random sequence population as shown in equations 4.2.3 through 4.2.6:

$$z = \frac{(p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}} \tag{4.2.3}$$

$$\hat{p} = \frac{p_1 + p_2}{n_1 + n_2} \tag{4.2.4}$$

Conditions: $$n_1 p_1 \geq 5, n_1(1-p_1) \geq 5 \tag{4.2.5}$$

$$n_2 p_2 \geq 5, n_2(1-p_2) \geq 5 \tag{4.2.6}$$

in which p_1 and p_2 are the proportions of SSE in populations 1 and 2 respectively, n_1 and n_2 are number of samples 1 and 2 respectively, and z is the z score. Assumptions for running a 2 proportion Z-test are the following: in each sample set, there must be at least 5 samples of the positive test *np* and 5 samples of the negative test n*(1-p)* at each location. Thus, in essence a SSE at a specific position cannot be completely conserved or observed in low number in either the aptamer or the random population sequence if it is to be compared with the 2 proportion Z-test.

4.2.7 SSE Base Position Dependent Mapping

For either the dominant secondary structures only case or both the dominant + suboptimal secondary structures, each SS$, both the random sequence population or the aptamer sequences are aligned 5' to 3' without any inserted gaps. At each base position the statistical frequency (or position frequency) of a particular SSE is determined by summing up the number of a particular SSE assignment at a particular base position over all sequences surveyed and dividing by the total number of sequences surveyed. To express this relationship as a position frequency matrix, one uses the following equations:

$$f_{k,j} = \frac{1}{N}\sum_{i=1}^{N} I(X_{i,j} = k) \tag{4.2.7}$$

$$i \in (1, \dots, N), j \in (1, \dots, l)\ k \in (S, H, L, I, G, D, or\ M)$$

In which *i* is a sequence number between 1 and N; $X_{i,j}$ is the number of samples with an SSE of *X* at base position *j*; *N* is the number sequences of base length *l*; *k* represents a given

SSE(S, H, L ,I, G, D, or M), and I is an indicator function where $I(X=k)$ is 1 if $X=k$, otherwise it is 0.

To illustrate the analytical outcome for a small sample size (N=4), an example schematic of how this analysis would be performed with for primary DNA structure analysis for 4 sequences is shown in Figure 4.2.6(B-C). This analysis starts with an input of unique sequences and generation of a position matrix and then, computation of a position frequency matrix.

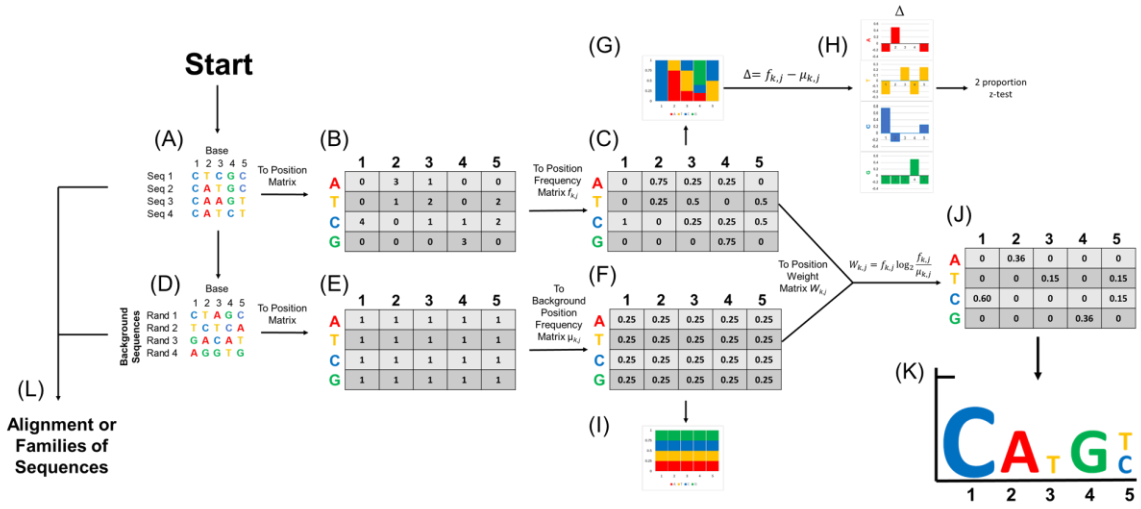


Figure 4.2.6. An example schematic to illustrate analytical methods used. For simplicity, analysis is first illustrated using primary structures of 4 sequences (of interest) against 4 random sequences (as a background). The steps include (A) input of sequences of interest, (B) representative position matrix of the shown sequences, (C) conversion of the position matrix into a position statistical frequency matrix ($f_{k,j}$). To include a background, analogous steps include (D) input of random sequences; and (E) generation of a position matrix of the random sequences; and (F) conversion of the position matrix into a position frequency matrix ($\mu_{k,j}$). Once $f_{k,j}$ and $\mu_{k,j}$ are generated, the first analytical pathway is to generate stacked bar graphs to plot the overall distribution of G, C, A and T for (G) sequences of interest, and (H) determine the numerical differences of positional frequency matrix for sequences of interest, $f_{k,j}$, from the random sequences positional frequency, $\mu_{k,j}$, ($\Delta = f_{k,j} - \mu_{k,j}$) with a one-tailed 2 proportion Z-test. The random sequences positional frequency can also be graphed as shown in (I) in which position frequency matrix, $\mu_{k,j}$, is graphed as a

stacked bar graph. (J) The second analytical pathway is to generate a position weight matrix (PWM), $W_{k,j}$, in which, ($W_{k,j} = f_{k,j} \log_2 \frac{f_{k,j}}{\mu_{k,j}}$), to then generate (K) a representative logo to visually indicate the relative significance of bases at a base position from a position specific score matrix (PSSM). Separate analytical pathways are shown in (L) in which analysis methods such as multiple sequence alignment are available.

Since additional variables, namely 7 SSE (rather than 4 bases), are considered in the secondary structure analysis, this same analysis must be expanded as illustrated in Figure 4.2.7

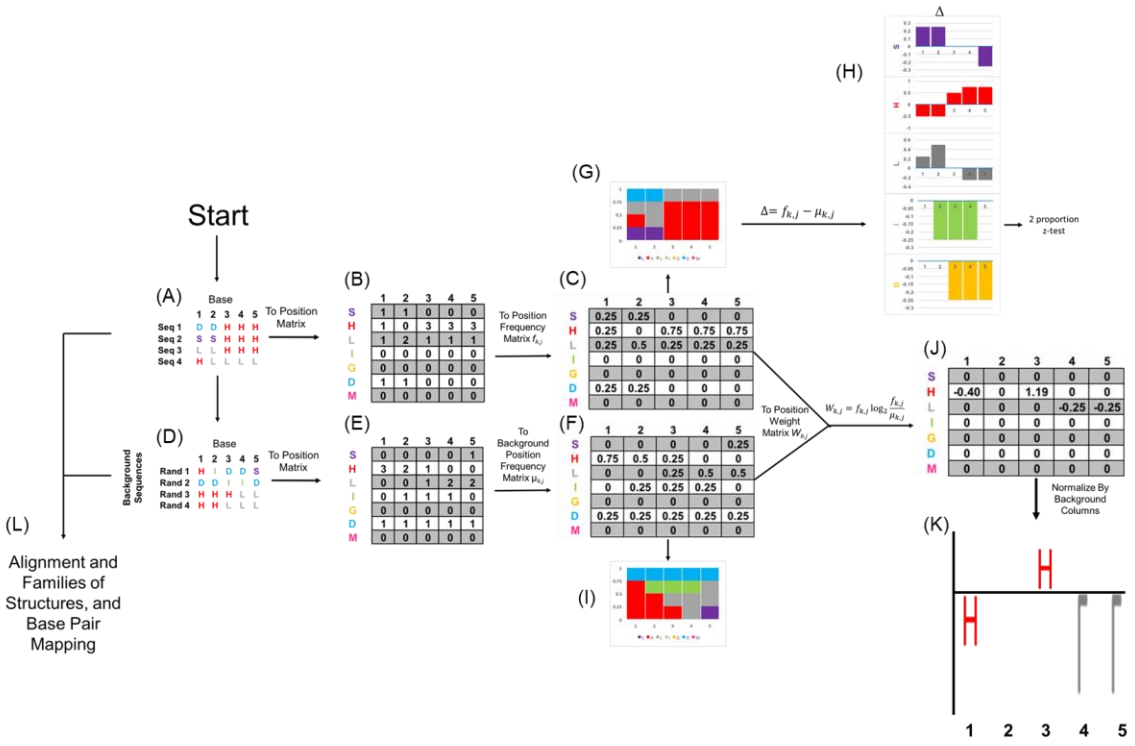


Figure 4.2.7 A example schematic to illustrate methods of analysis used with secondary structure strings (SS$s). The steps of the analysis are (A) input of aptamer SS$s, (B) generation of a position matrix of the shown SS$s; and (C) conversion of the position matrix into a position statistical frequency matrix ($f_{k,j}$). To include a background analogous steps include (D) input of random sequence population SS$s; (E) generation of a position matrix of the random sequence SS$s; and (F) conversion of the position matrix into the position frequency matrix ($\mu_{k,j}$). Once $f_{k,j}$ and $\mu_{k,j}$ are generated, the first analytical pathway is to generate stacked bar graphs to plot the overall distribution of the 7 SSE for (G) the aptamer sequences. (H) The difference can be determined between the aptamer sequences and the random sequence population from Δ, in which $\Delta = f_{k,j} - \mu_{k,j}$,with a 2 proportion Z-

test (I) It is also possible to create a stacked bar graph of the position frequency matrix, $\mu_{k,j}$. (J) The second analytical pathway is to generate a position weight matrix $W_{k,j}$ in which , $W_{k,j} = f_{k,j} \log_2 \frac{f_{k,j}}{\mu_{k,j}}$ to then generate (K) a representative logo to visually indicate significance of a SSE at a particular position. This logo serves as a position specific score matrix (PSSM). Separate pathways for analysis are also available such as (L) base pair mapping or alignment.

After generating the position statistical frequency matrix $f_{k,j}$ for aptamers and the $\mu_{k,j}$ for the random sequences, the statistical frequency for each SSE is plotted as a function of base position j in a stacked bar graph for both the aptamer and the separate random sequence population for two cases of dominant secondary structures only or for the case including both dominant and suboptimal secondary structures. Graphing the position frequency matrix allows for a "fragmentation map" of secondary structure elements to be visualized across all sequences for both the aptamers and the random sequence population. The numerical differences in position frequency matrices for the aptamer and the random sequence population along with one-tailed 2 proportion Z-tests (see equation 4.2.3 through equation 4.2.6) to determine any statistical significance in Δ values at each of the 69 base positions..

4.2.8 Computation of Position Specific Score Matrices (PSSM)

While the position-dependent Δ chart takes into account numerical differences between SSE in aptamers versus a random sequence background, it does not factor in potential dependence of an SSE at a base positions proportion on the subpopulation of other SSE at that base position. The next set of analyses, to compute a position specific score matrix (PSSM), accounts for amount of disorder or entropy at a position in the background compares the entropy in the sample as a signal to noise ratio. This way of describing the

difference between an aptamer and the background allows better significance assessment for observed changes.

To compute a PSSM, first a position matrix (P) was made from the aptamer sequences as shown in Figure 4.2.7(A-B) and described by equation 4.2.8:

$$P_{k,j} = \sum_{i=1}^{N} I(X_{i,j} = k) \tag{4.2.8}$$

$$i \in (1, \dots, N), j \in (1, \dots, l), k \in (S, H, L, I, G, D, M)$$

Then a background position frequency matrix ($\mu_{k,j}$) is determined with either the dominant secondary structure only random sequence population or with the dominant + suboptimal SS$s of the random sequence population to use as a background against the corresponding aptamer SS$s as illustrated in Figure 4.2.7(E-F) using the following in equation:

$$\mu_{k,j} = \frac{1}{N} \sum_{i=1}^{N} I(X_{i,j} = k) \tag{4.2.9}$$

$$i \in (1, \dots, N), j \in (1, \dots, l), k \in (S, H, L, I, G, D, M)$$

To add the possibility to perform additive smoothing to aptamers, a pseudocount function for aptamer SSE statistical frequency versus a known background (generated random sequence population), was created as shown in the following equation

$$\theta_{k,j} = \frac{X_j + 7\mu_{k,j}\alpha}{N + 7\alpha} \tag{4.2.10}$$

in which X*j* is the number of samples with an SSE of X=k at base position *j*; N is the number sequences of length l, $\mu_{k,j}$ is the random sequence population background incidence rate (or statistical frequency) of an SSE represented by k at base position *j*, α is the additive smoothing parameter, and 7 is the total number of SSE possible (with each hairpin split into its stem and loop components), and thus corresponds to the size of our alphabet k. The pseudocount frequency function is then used to create a normalized position probability matrix with alpha values tested ranging from 0-2 with the equation:

$$\theta_{k,j} = \frac{7\alpha\mu_{k,j} + \sum_{i=1}^{N} I((X_{ij} = k))}{7\alpha + N} \tag{4.2.11}$$

$$i \in (1, \dots, N), j \in (1, \dots, l), k \in (S, H, L, I, G, D, M)$$

Finally, a Position Weight Matrix (PWM) ($W_{k,j}$) was created similarly as shown in Figure 4.2.7(J) and is generated with the following equation:

$$W_{k,j} = \frac{1}{N}\log_2\left(\frac{\theta_{k,j}}{\mu_{k,j}}\right)\sum_{i=1}^{N} I(X_{ij} = k) \tag{4.2.12}$$

$$i \in (1, \dots, N), j \in (1, \dots, l), k \in (S, H, L, I, G, D, M)$$

To create a position specific score matrix (PSSM) ($S_{k,j}$), the PWM has to be normalized to the amount of information content contained originally at each base position *j* in the background position frequency matrix, $\mu_{k,j}$ as shown in the Figure 4.2.7(K). This

normalization of height of the letter displayed in the PSSM to information content in each column j in the background frequency matrix $\mu_{k,j}$ was expressed as:

$$H_j = \frac{\log_2(7)}{\sum_{columns\ j} \log_2(\frac{1}{\mu_{k,j}})(\frac{1}{N}) \sum_{i=1}^{N}(X_{i,j} = k)} \tag{4.2.13}$$

$$i \in (1, \dots, N), j \in (1, \dots, l), k \in (S, H, L, I, G, D, M)$$

Thus, the overall expression for the position score matrix is (PSSM) ($S_{k,j}$),

$$S_{k,j} = H_j W_{k,j} \tag{4.2.14}$$

From this expression in equation 4.2.14, the sequence logo graphic corresponding to either the dominant or dominant + suboptimal set of SS$s was created in Matlab 2016.

4.2.9 Scoring Secondary Structure Strings (SS$s) with PSSM

Scoring SS$s is simply summing up the PSSM for a particular SS$ of a sequence i over each base position j:

$$\tau_i = \sum_{row\ i} \sum_{j=1}^{l} I(X_{i,j} = k) S_{k,j} \tag{4.2.15}$$

$$i \in (1, \dots, N), j \in (1, \dots, l), k \in (S, H, L, I, G, D, M)$$

Where τ_i is the score from a PSSM for each SS$ i , where j is the base position from 1 to length l, and $X_{i,j}$ is the SS$ with an SSE of X at base position j, k is the set of symbols for

SSE(S, H, L ,I, G, D, M), I is an indicator function where $I(a=k)$ is 1 if $a=k$, otherwise it is 0, and $S_{k,j}$ is the position specific score matrix.

4.2.10 Multiple Secondary Structure String Alignment (MSSA)

Four different alignments were performed on the SS$: aptamer dominant SS$s, generated random sequence population dominant SS$s, aptamer dominant + suboptimal SS$s, and generated random sequence population dominant + suboptimal SS$s. SS$s alignment is performed with CLUSTALW in MEGA7[333] with SS$s that using amino acid substitutions shown in Table 4.2.1. In essence, this adaptation allows for the use of existing protein sequence alignment programs to now align secondary structure in DNA aptamers from SS$s. To prevent artifacts due to intrinsic assumption of various amino acid substitutions used (e.g. substituting aspartic acid with glutamic acid since both are acidic residues), an identity matrix for the substitution alignment matrix was used along with removal of residue specific penalties and hydrophilic penalties. All other settings, in MEGA7 for protein alignment in CLUSTALW were the default settings in the software program at the time data analysis was carried out in version 7.0.26 from 7/4-7/25 2017. Data was exported into Matlab 2016 for analysis by developed program algorithms, and output to Microsoft Excel for graphing purposes.

4.2.11 Identified Aligned SS$ Domain Analysis

After alignment of the SS$, continuous regions of alignment with greater than 50% consensus were called consensus domains. Regions of the alignment containing greater than 50% inserted gaps were simply called gap regions (*G)*. Regions containing a majority of SSEs but not having consensus were assigned the identity of non-gaps (*N*). After

covering these possible categories of alignment regions, to display large data sets succinctly, representations of an overall alignment were created, so summary data for each region and overall alignment could be easily displayed. Data displayed includes: average length and standard deviation of length for each region in the alignment; the average number of consensus domains observed per sequence; the average percentage of SS$s associated with domains. For consensus domains, additional information is provided: the associated SS$ of the domain, average consensus across a domain (Conserved), statistical frequency of observation of a domain in sequence set (Frequency), and fraction of the domain associated with the fixed base regions (Fraction Fixed Base).

4.2.12 Generation of Distribution of Consensus Domains

After alignment to identify consensus domains, gaps were removed from the alignment so base position dependence of the SS$ could be investigated in terms of the identified consensus domains. Using algorithms developed in Matlab 2016 these SS$ domains were analysed for their base position specific distribution; and base position distance between identified domains: average distance from an identified domain n to domain n+1 ($d_{(n,n+1)}$), average distance from a domain to the closest previously occurring domain or SS$ end 5' to 3' (d_{prior}), average distance from a domain to the next occurring domain or SS$ end 5' to 3' (d_{next}); and average number of bases lost from the central portion of a domain (domain loss). Plots of the distributions of the identified domain distributions were created in Microsoft Excel 2016.

4.3 Results and Discussion

4.3.1 Analysis of Random Sequence Population

4.3.1.1 Analysis of Distribution of Base Identities at Each Base Position

Analysis of the random region at each of the 40 base positions showed an average maximum consensus of a single base at a position of 26.6% with a standard deviation of 0.76%, the average minimum consensus at a base position of 23.4% with a standard deviation of 0.80%, maximum consensus of a base at any base position of 28.6%, and minimum consensus of a base at any base position of 21.1% indicating an approximate equibase distribution of 25% in the 40 base long central segment of the generated random sequence population.

4.3.1.2 Analysis of MSA of Generated random sequence population Random Segment

Multiple sequence alignment (MSA) of the central segment by MUSCLE in MEGA7 shows a maximum consensus of 42.8% at one base position and only 3 other positions having a maximum consensus with values at or above 40%. Using this analysis tool, the random sequence population aligned primary structure does not indicate clear structural patterns.

4.3.2 Base-Pairing Mapping Analysis of Aptamers

In base pairing heat maps for aptamers sequence information is removed and only base pairs overlapping with other base pairs from separate secondary structures of aptamers are considered. In Figure 4.3.1, only dominant secondary structures are considered and indicates a large occurrence of overlap with 25 of the 42 structures having overlap at base positions (63, 59) and (64, 58) and 21 of the 42 structures having overlaps at base positions (65, 57), (66, 56), and (67, 55). While this could possibly be considered significant, it is

notable that all these pairings occur in the strictly fixed base region, and thus involve identical base identities and result in identical self-folding. The next most common occurrence of overlapped base pairings in dominant structures occurs at base position (31, 14) with 4 of the 42 dominant structures having overlap at that point. In contrast to the prior overlaps, however, base 31 is within the central random region.

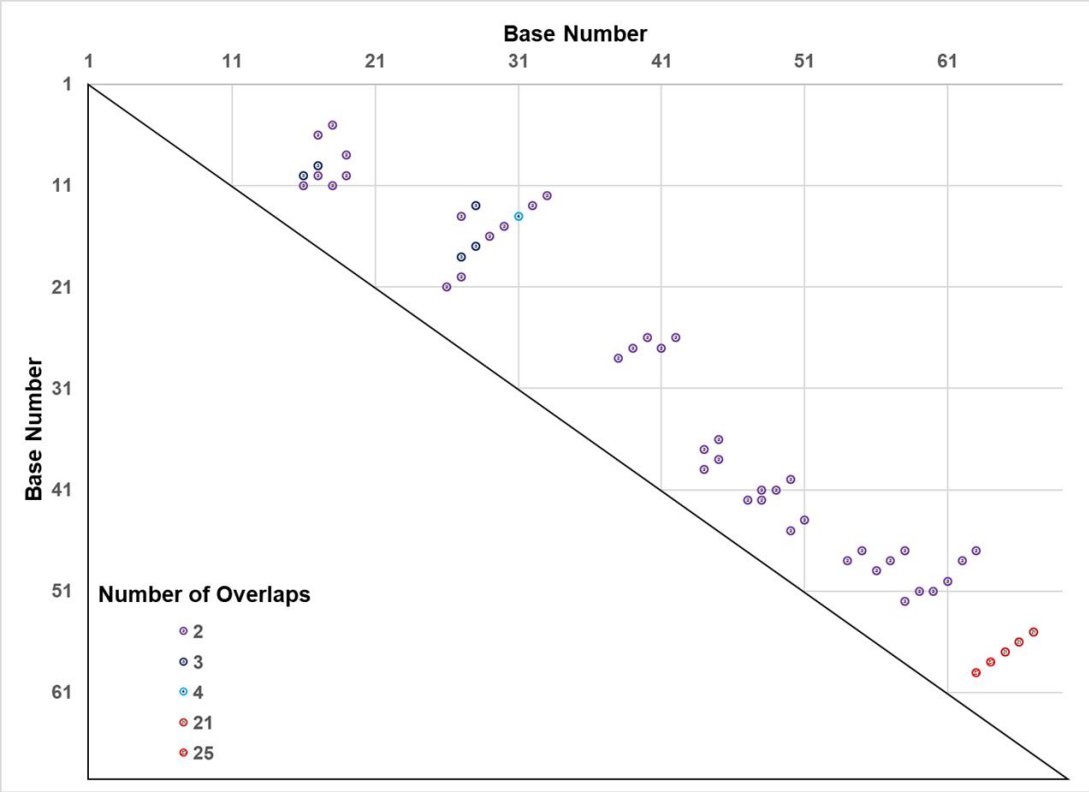


Figure 4.3.1. Heat map to show total number of all shared, position-dependent Watson Crick (A-T, T-A, G-C, or C-G) or Wobble (G-T; T-G) base-pair interactions for two or more dominant predicted secondary structures for all AuNR aptamer sequences.

The inclusion of suboptimal secondary structures with the dominant secondary structures is examined in Figure 4.3.2. Similar base pairing overlap behavior was observed at base positions (63, 59) and (64, 58) and for (65, 57), (66, 56), and (67, 55) with the same difference of 4 overlaps between the two sets of pairings. More occurrences of base overlap

is generally observed with the inclusion of the suboptimal structures, but it should be noted that normally a single sequence is more likely to contain identical base overlap sections in both its dominant and suboptimal structures resulting in "intra-sequence" base pairing overlap, rather than the possibly less common "inter-sequence" base pairing overlap between different aptamer sequences. Further analysis of base pairing will distinguish sequence and structure identity to enable further interpretation of base pair overlap in section 4.3.4.

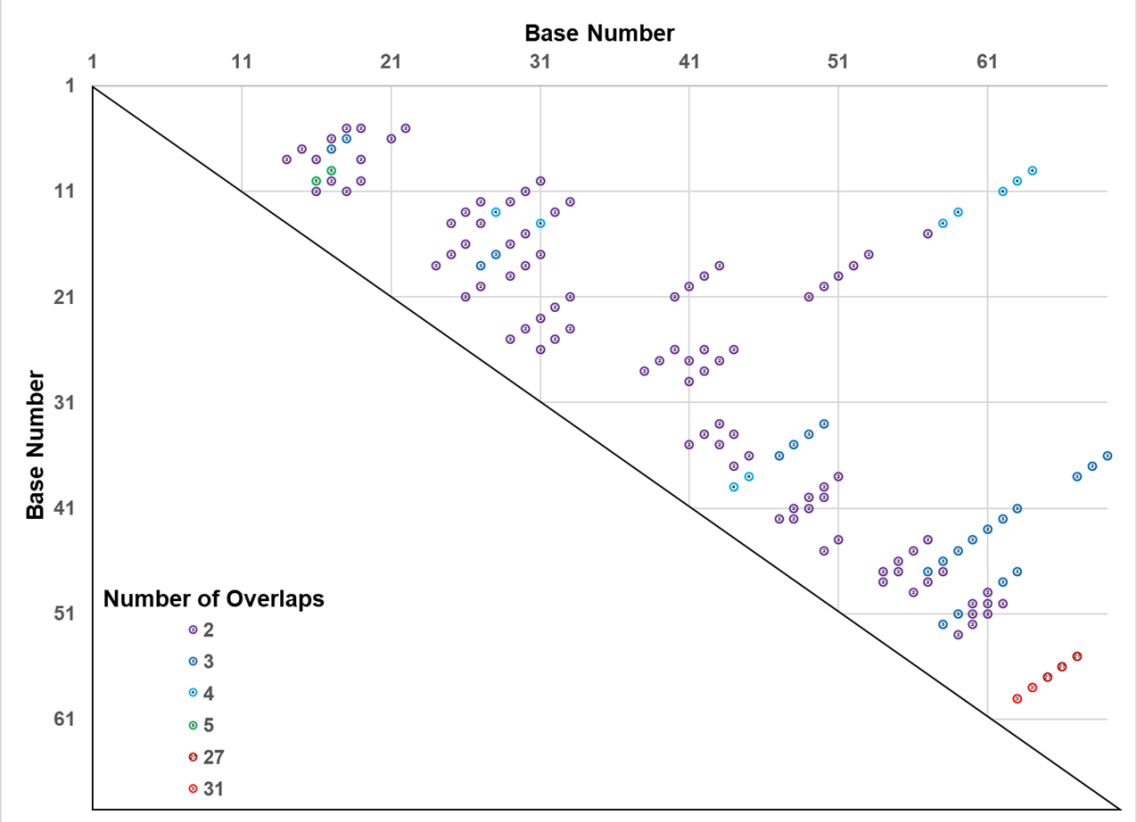


Figure 4.3.2. Heat map to show total number of all shared, position-dependent Watson Crick (A-T, T-A, G-C, or C-G) or Wobble (G-T; T-G) base-pair interactions for two or more dominant and/or suboptimal predicted secondary structures for all AuNR aptamer sequences

4.3.3 Secondary Structure Family (SSF) Analysis

Using programs in Matlab 2016 for equations shown in Table 4.2.2 through, eleven secondary structure families (SSF) emerge among AuNR aptamers that possess more than one member, as listed in Table 4.3.1. SSF9 has the most (10) members, consisting of 8 dominant structures and 2 suboptimal aptamer structures.

Table 4.3.1. List of eleven Secondary Structure Families (SSFs) with at least two members, along with corresponding information on the total number of each secondary structure element (SSE), aptamer nomenclatures and number of members as well as the numbers of dominant and suboptimal secondary structures in each family among all AuNR aptamers identified via CompELS

Family No.	Secondary Structure Family (SSF)	Aptamer Nomenclature and Designation as Dominant (*.S1) or Suboptimal (*.S2, and *.S3) Secondary Structure of All Family Members Identical sequences in the same family are underlined	No. of Family Members	No. of Dominant (Suboptimal) Secondary Structures
1	2SS 2HP 1IL 0B 1DS 0MB	111.S1, 116.S1, 116.S2	3	2 (1)
2	2SS 2HP 1IL 1B 2DS 0MB	101.S1, 117.S1, 418.S1	3	3 (0)
3	3SS 2HP 0IL 0B 0DS 0MB	111.S2, 111.S3, 120.S1, 214.S1, 403.S1, 412.S2	6	3 (3)
4	3SS 2HP 0IL 1B 1DS 0MB	401.S1, 420.S1	2	2 (0)
5	3SS 2HP 1IL 0B 1DS 0MB	108.S2, 112.S1, 406.S1	3	2 (1)
6	3SS 2HP 2IL 0B 2DS 0MB	110.S1, 115.S1, 416.S1	3	3 (0)
7	4SS 3HP 0IL 0B 0DS 0MB	102.S1, 212.S1, 217.S1, 219.S1, 404.S1, 412.S1	6	6 (0)
8	4SS 3HP 0IL 1B 1DS 0MB	101.S2, 118.S1, 201.S2, 210.S1	4	2 (2)
9	4SS 3HP 1IL 0B 1DS 0MB	105.S1, 108.S1, 114.S1, 119.S1, 204.S1, 204.S2, 402.S1, 413.S1, 415.S2, 421.S1	10	8 (2)
10	5SS 4HP 0IL 0B 0DS 0MB	217.S2, 410.S1, 417.S1, 419.S1	4	3 (1)
11	5SS 4HP 1IL 0B 1DS 0MB	409.S1, 414.S1	2	2 (0)

The same analysis for SSF was performed on the generated random sequence population secondary structures, so they could be used as a background for the analysis of the SSF of the aptamers. The most commonly observed SSFs for the random sequence population in the dominant structures were the (4,3,0,0,0,0) family with 131 members out of 1000

structures and (4,3,1,0,1,0) family with 85 members out of 1000 structures. When both the dominant and suboptimal secondary structure cases are considered for SSFs, the (4,3,0,0,0,0) family and (4,3,1,0,1,0) family still predominate in the random sequence population with 197 members out of 1673 structures and 136 members out of 1673 structures, respectively. Several additional 6-coordinate values emerge though some have only 1 sequence member and thus do not constitute a SSF This background of random sequence population SSFs was used for the one tailed 2-proportion Z-test to attempt to identify families that occurred at a significantly higher rate than the background as shown in Figure 4.3.3 for the dominant secondary structures and Figure 4.3.4 for the dominant + suboptimal secondary structures.

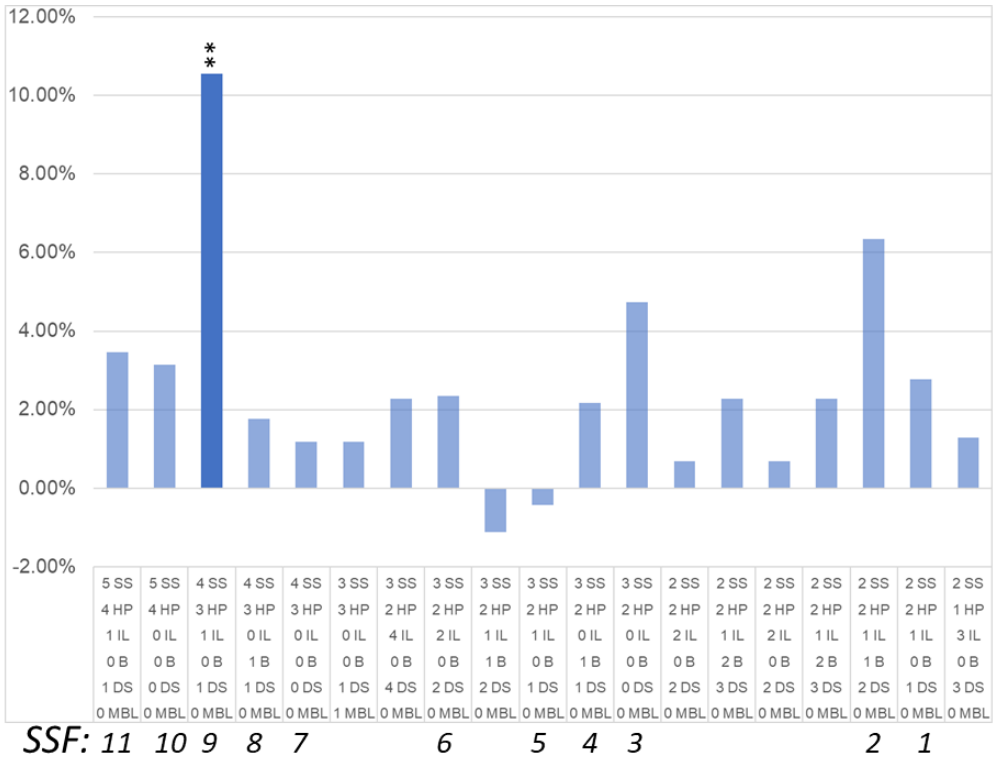


Figure 4.3.3. Differences (%) of 6-coordinate values (displayed vertically) for SSE in aptamer sequences versus the random sequence population for only dominant secondary

structures. Where relevant, aptamer SSF number identities are listed under the appropriate 6-coordinate column. ** indicates 97.5% CI for significant difference in proportion between the dominant aptamer secondary structures in SSF9 and that of the dominant random sequence population secondary structures from a one tailed 2 proportion Z-test.

From the Z-test analysis comparing proportion of the dominant secondary structures, the differences in the proportion of the aptamers (SSF9) and the random sequence population with 4 single strand segments, 3 hairpins, 1 internal loop, and 1 duplex or (4, 3, 1, 0, 1, 0) appears significant. SSF9 members are shown in Figure 4.3.5 (9-10). It should be noted however that this analysis, while statistically valid, is performed with low sample numbers.

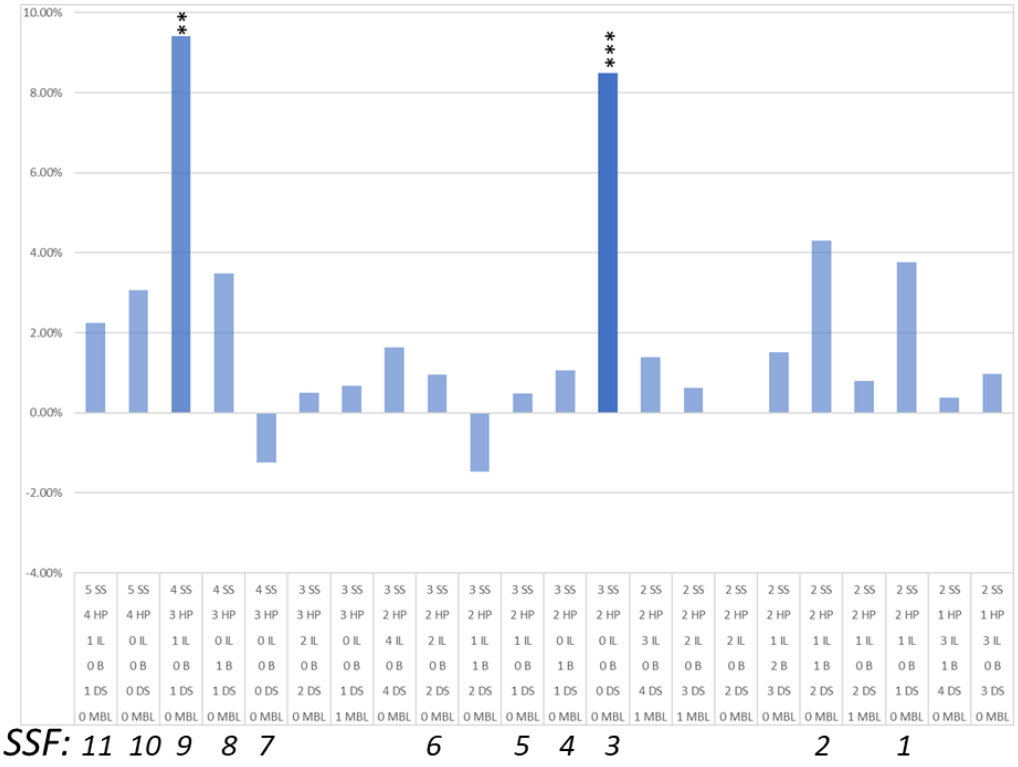


Figure 4.3.4 Differences (%) of 6-coordinate values (displayed vertically) for SSE in aptamer sequences versus the random sequence population for dominant + suboptimal secondary structures. Where relevant, aptamer SSF number identities are listed under the appropriate 6-coordinate column. ** and *** indicates 97.5% and 99.5% CI for significant difference in proportion between the dominant + suboptimal aptamer secondary structures

in SSF9 and SSF3 to that of the dominant random sequence population secondary structures from a one tailed 2 proportion Z-test.

By including consideration for suboptimal secondary structures in Figure 4.3.4, aptamer secondary structures in SSF3 with 3 single-stranded segments and 2 hairpins or (3, 2, 0, 0, 0, 0) also emerges as statistically significant. SSF3s' members are shown in Figure 4.3.5 (3).

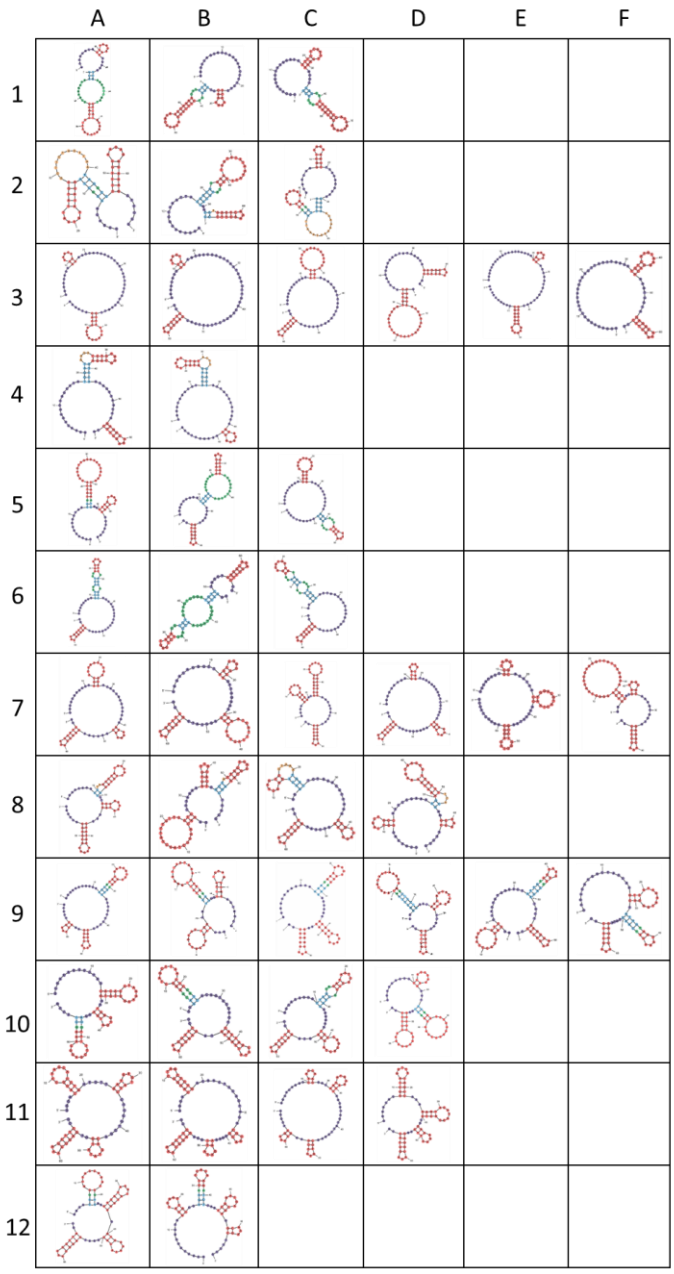


Figure 4.3.5. All Secondary Structure Families (SSF) by row (1) SSF1: (2,2,1,0,1,0): (2)SSF2: (2,2,1,1,2,0), (3) SSF3: (3,2,0,0,0,0), (4) SSF4: (3,2,0,1,1,0), (5) SSF5: (3,2,1,0,1,0), (6) SSF6: (3,2,2,0,2,0), (7) SSF7: (4,3,0,0,0,0), (8) SSF8: (4,3,0,1,1,0), (9-10) SSF9: (4,3,1,0,1,0), (11) SSF10: (5,4,0,0,0,0), (12) SSF11: (5,4,1,0,1,0). Individual secondary structures by position: (A1) 111.S1, (B1) 116.S2, (C1) 116.S2, (A2) 101.S1, (B2) 117.S1, (C2) 418.S1, (A3) 111.S2, (B3) 111.S3, (C3) 120.S1, (D3) 214.S1, (F3) 403.S1, (G3) 412.S2, (A4) 401.S1, (B4) 420.S1, (A5) 102.S2, (B5) 112.S1, (C5) 406.S1, (A6) 110.S1, (B6) 115.S1, (C6) 416.S1, (A7) 102.S1, (B7) 212.S1, (C7) 217.S1, (D7) 219.S1, (E7) 404.S1, (F7) 412.S1, (A8) 101.S2, (B8) 118.S1, (C8) 201.S2, (D8) 210.S1,

(A9) 105.S1, (B9) 108.S1, (C9) 114.S1, (D9) 119.S1, (E9) 204.S1, (F9) 204.S2, (A10) 402.S1, (B10) 413.S1, (C10) 415.S2, (D10) 421.S1, (A11) 217.S2, (B11) 410.S1, (C11) 417.S1, (D11) 419.S1, (A12) 409.S1, (B12) 414.S1

4.3.4 Base Pairing with Individual Aptamer Secondary Structure and SSF

Continuing with base pairing analysis from section 4.3.2, base pairing for each aptamer sequences (each with a unique symbol) are plotted to identify overlap within identified secondary structure families (SSF) only (intra-family pairings) and for overlap between and within SSF (intra- + inter-family pairings). Figure 4.3.6 shows this mapping for dominant secondary structure only and is thus directly comparable to the heat map shown in Figure 4.3.1. Base segments with either intra-family or intra- + inter-family base-pairing are circled and SSF identities are included. Immediately apparent is the intra-family pairing of SSF9 at the previously identified location of (31,14). The entire secondary structures for the aptamer sequences included in SSF9 involved in this base-pair overlap are shown in Figure 4.3.8 with bases in this overlapping base pairing highlighted. Additionally, in this dominant secondary structure base pairing map in Figure 4.3.6, base segments with

numerous overlapping base pairings in bases 63-67 consist of intra- and inter-family within and between SSF9 and SSF3.

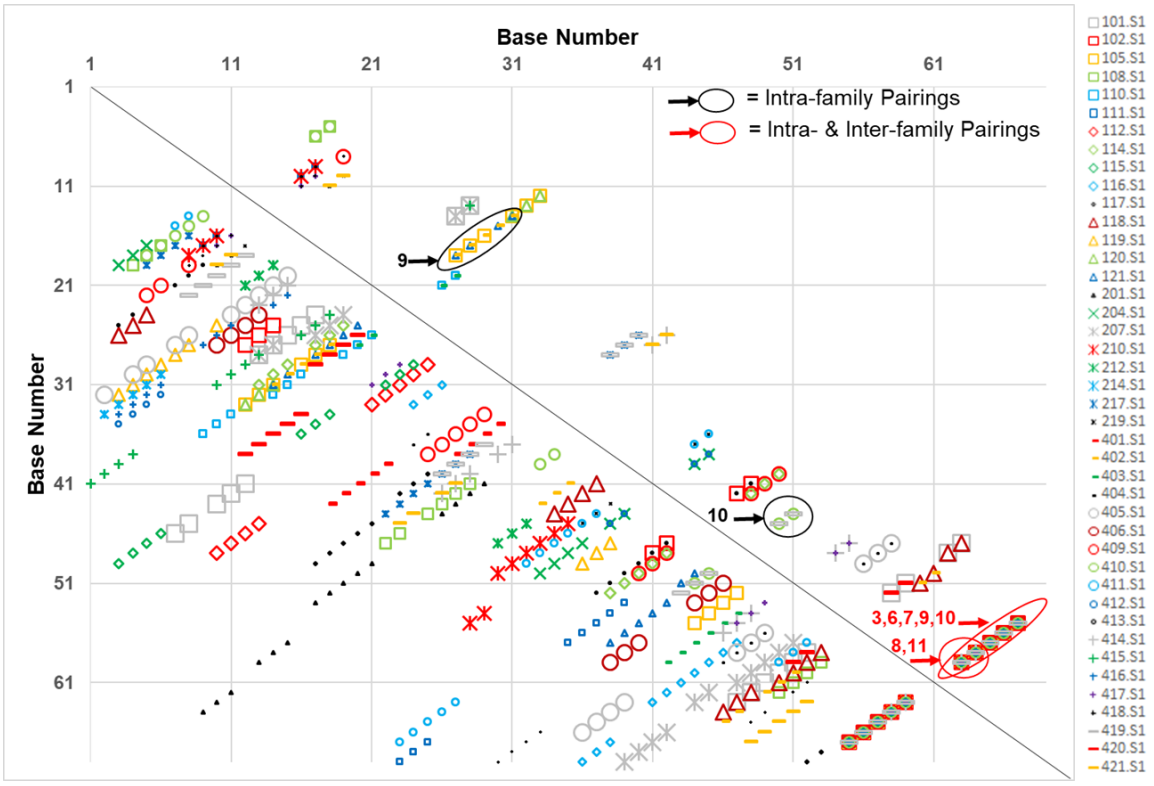


Figure 4.3.6. Base pair map indicating all position-dependent base-pair interactions for dominant predicted secondary structure for each aptamer sequence (lower left half) and exclusively shared (i.e. between two or more aptamer sequences) position-dependent base-pair interactions (upper right half). Where relevant, shared intra-family and shared intra- + inter-family pairings are circled and the number(s) of the SSF(s) is provided. A legend of symbols for all dominant secondary structures of each AuNR aptamer mapped is provided.

When the base pair mapping is extended to include dominant + suboptimal structures in Figure 4.3.7, as we saw previously in Figure 4.3.2, the number of base-pair overlaps is dramatically increased. The same pattern observed for intra- and inter-family overlap in the dominant base pairing map at bases 63-67 (Figure 4.3.6) is repeated in Figure 4.3.7 for dominant + suboptimal base pairing; however, many of the additional pairings are observed to arise from intra-family pairings in SSF1 at bases 57 to 69 and in SSF 9 from

bases 44 to 51. Thus, while the heat maps in Figure 4.3.1 and Figure 4.3.2 proved purely numerical information of base-pairings, these maps in Figure 4.3.6 and Figure 4.3.7 provide the identity of aptamer sequence and SSF involved.

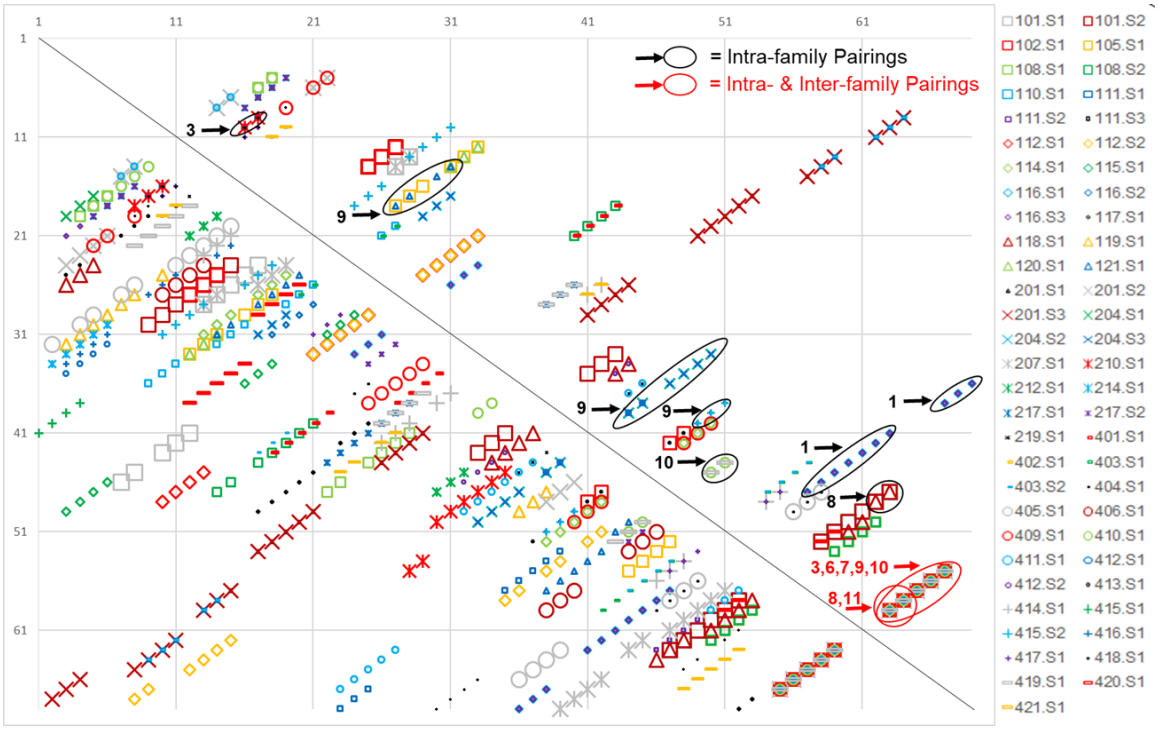


Figure 4.3.7. Base pair map indicating all position-dependent base-pair interactions for dominant + suboptimal predicted secondary structure for each aptamer sequence (lower left half) and exclusively shared (i.e. between two or more aptamer sequences) position-dependent base-pair interactions (upper right half). Where relevant, shared intra-family and shared intra- + inter-family pairings are circled and the number(s) of the SSF(s) is provided. A legend of symbols for all dominant + suboptimal secondary structures of each AuNR aptamer mapped is provided.

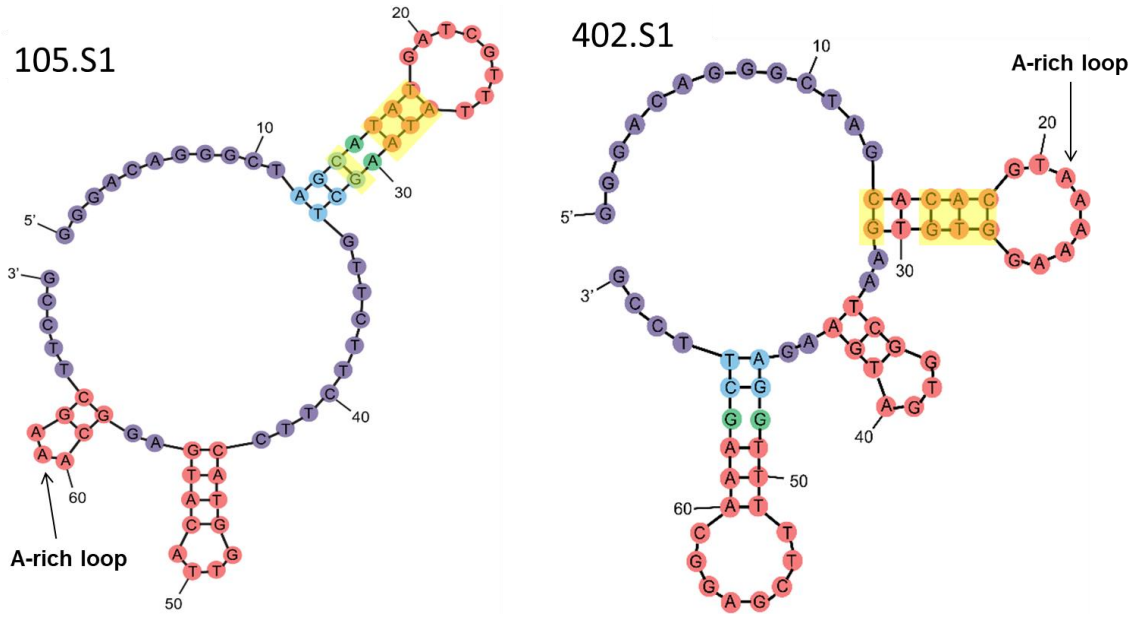


Figure 4.3.8. Two predicted secondary structures color-coded to indicate shared and unshared base-pairs indicated in the adjacent base-pair map for 105.S1 and 402.S1 that both belong to SSF 9. Shared base pairs at base (14, 31) and, (16-18, 29-27) are highlighted. Hairpins with A-rich loops are also indicated.

4.3.5 Analysis of Mapping of Secondary Structure Elements (SSE) at Each Base Position

The frequency of SSE at base position, or the matrix $f_{k,j}$ for aptamers or $\mu_{k,j}$ for the (background) random sequence population is shown in Figure 4.3.9. From examination of each case (A)-(D) among all four cases, that general trends in the distribution of SSE hold for all four cases. Shared trends in the base position dependent distribution of SSE include the following: (1) single-stranded segments are prevalent near and at the 5' and 3' ends; and (2) predominant existence of a hairpin loop structure at base positions 60-62 with the flanking hairpin stem at base positions 55-59 and 63-64. Distinction in the aptamers versus background also occur. For example, duplexes seem to be favoured more in the background structures around base position 9 and base position 64. Internal loops also seem prevalent in background at the same location as the hairpin loop at base positions 60-62. Single-

stranded segments in the aptamers appear to predominate at base positions 20, 35, and 54 within the fixed base segment. While a hairpin loop appears to predominate at base position 12 in the random sequences, appearance of a hairpin loop in the aptamer sequences is generally more prevalent and exhibits a broader distribution about base position 12. Another difference observed at base position 12 is the occurrence of an internal loop in the background that does not appear in the aptamer. Lastly, bulges and especially multibranch loops do not seem particularly favourable in either the random sequence population background or the aptamer sequences. Moreover, both SSE do occur at more base positions in the random sequences. In the dominant + suboptimal aptamer secondary structures (Figure 4.3.9(B)), the multibranched loop appears to peak at the same base position of an internal loop, and the conserved hairpin loop in base position 60-62.

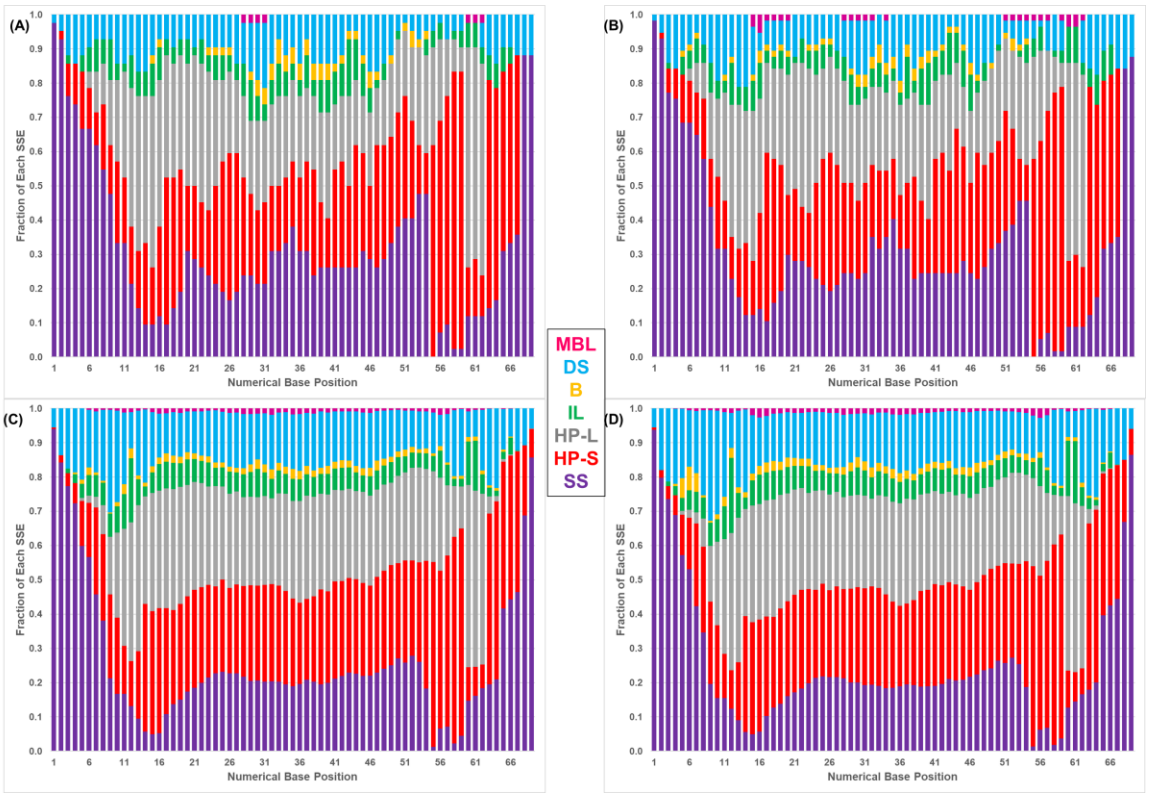


Figure 4.3.9. Distribution of secondary structure elements (SSE) as a function of numerical base position for self-hybridized (A) dominant structures of the aptamers (B) dominant + suboptimal structures of the aptamers (B) dominant structures of the generated random sequence population background and (D) dominant + suboptimal structures of the generated random sequence population background. Each bar is color-coded to indicate the statistical frequency of single-stranded segments (purple), hairpin stems (red); hairpin loops (gray); interior loops (green); bulges (yellow); duplexes (blue); and multibranched loops (dark pink) at a given base position.

To facilitate further quantitative interpretation of the distribution of secondary structure elements (SSE), the numerical percent difference in position-dependent distribution of SSE between the aptamers and the random sequence background (for either dominant or dominant + suboptimal structures) are calculated and one-tailed two proportion Z-tests are performed where suitable. These quantitative results are shown in Figure 4.3.10 and Figure 4.3.11.

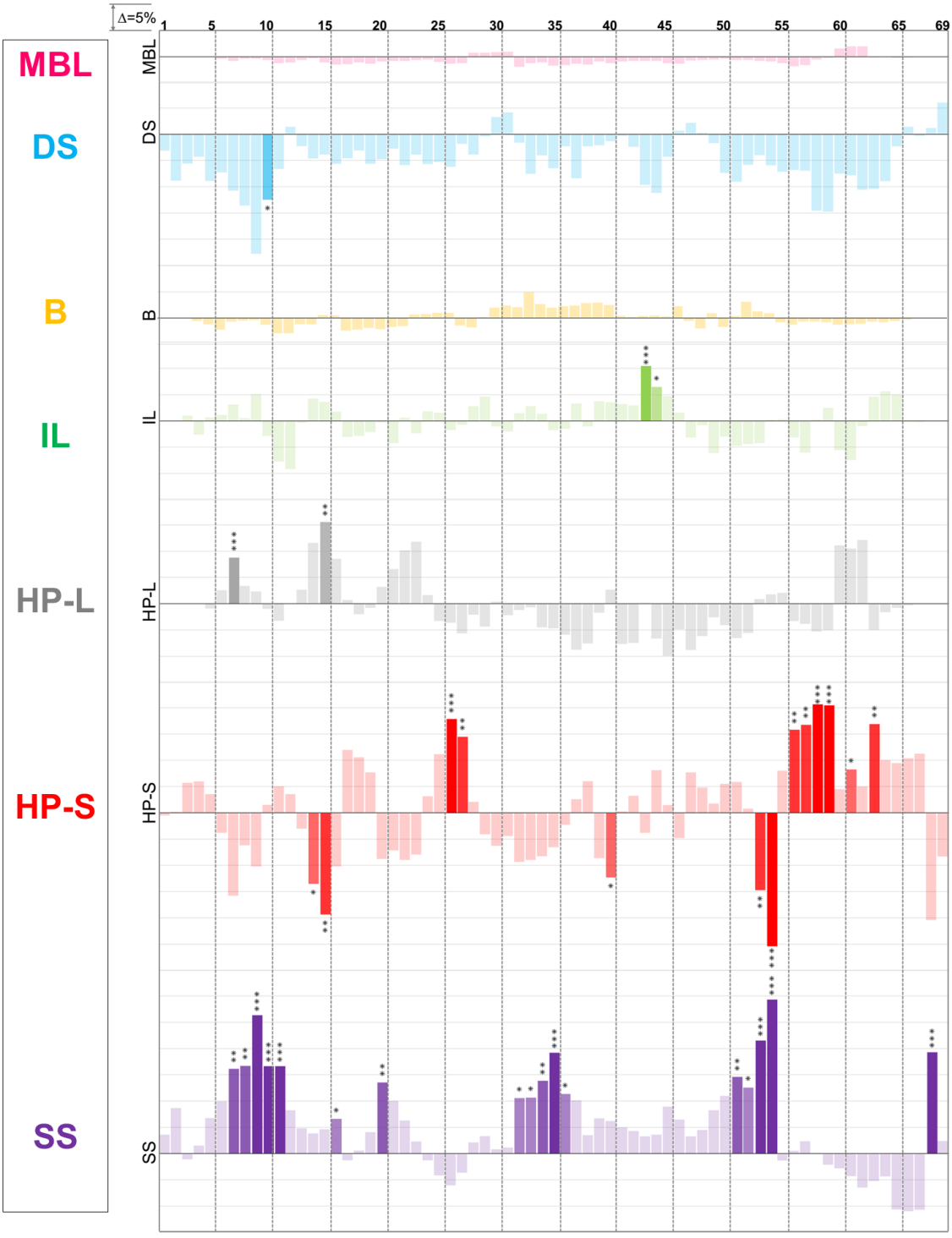


Figure 4.3.10. ΔSSE (%): Percentage difference in dominant SSE at each base position between aptamer and random sequence populations. *, **, and *** series denote a 95%,

97.5%, and 99.5% CI, respectively for significant differences based on a one-tailed 2 proportion Z-test.

For the dominant only structures case in Figure 4.3.10, several base positions present statistically significant differences in SSE distribution between the aptamer and the random sequence population. Duplex segments at base position 10 and hairpin stems at base positions 14-15, 40, and 53-54 occur less in the aptamers. The decrease in duplex segments at position 10 appear offset, however, by a significant increase in single-stranded segments from base number 7-11. Likewise, the decrease in the hairpin stems at base positions 14-15 is associated with an increase in hairpin loop formation at base position 15; and decrease in the hairpin stems at base positions 53-54 are associated with an increase in single-stranded segments at base positions 51-54. In tying these SSE-specific results back to the base-pair map shown in Figure 4.3.1 in dominant structures, the significant increase in the occurrence of hairpin stems in the aptamer at base position 56-59 accounts for the highest number of observed base pair mapping overlap at those base positions. More single-stranded segments appeared to be in the aptamer in the both the central random region at base positions 32-36 as well as near 5' and 3' ends. Internal loops, only at base positions 43-44, exhibit significant differences in occurrence.

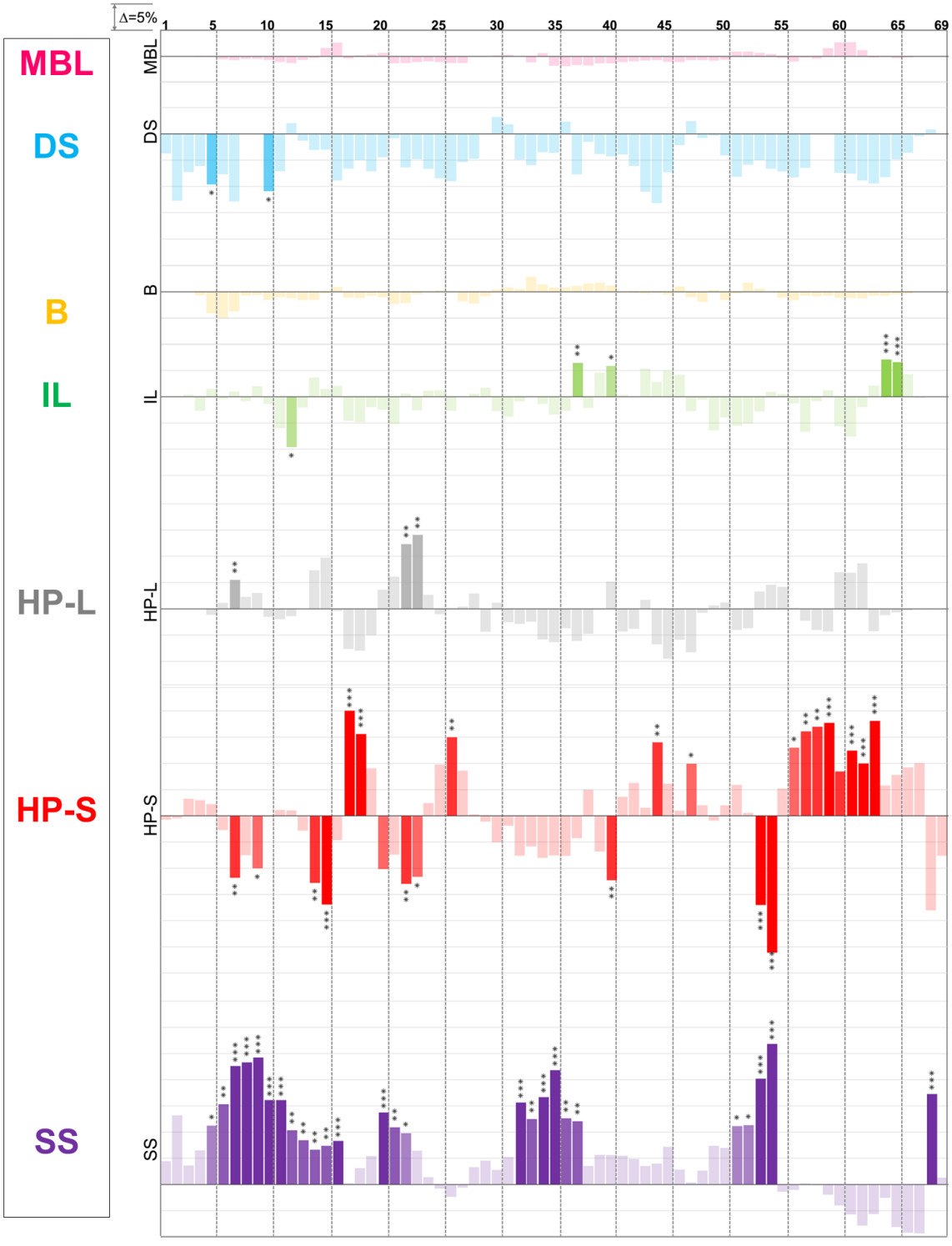


Figure 4.3.11. ΔSSE (%): Percentage difference in dominant + suboptimal SSE at each base position between aptamer and random sequence populations. *, **, and *** series

denote a 95%, 97.5%, and 99.5% CI, respectively for significant differences based on a one-tailed 2 proportion Z-test

For SSE distribution from both the dominant + suboptimal structures case in Figure 4.3.11, similar trends as the dominant structures are observed. Increase in single-stranded segments is observed in base position 5-16, 20-23, 32-37, 51-54, and 68. The increase of single-stranded at base position 7,9,14-15, 20, 22, and 53-54 appears to be counterbalanced by a decrease in the number of hairpin stems while the increase in single-stranded segments at base positions 7 and 22 appear associated with an increase in hairpin loop formation. Significant increase in hairpin stem formation from base positions 56-63 is similar to the observed same trend for the dominant only structures in Figure 4.3.10; however, interestingly, this increase in hairpin stems at these base positions does not correspond to any significant increases in hairpin loops. The decrease in duplexes near the 5' end is also comparable to that of the dominant only case. Finally, internal loop formation appears more prominent in both the central random region and near the 3' end.

After examining strictly numerical differences in SSE occurrences at each base position with just the difference in the aptamer compared to the random sequence population, a more “weighted” approach was untaken to compute a position specific score matrix (PSSM) from the same data for the case of the dominant only structures as shown in Figure 4.3.12, as well as for the dominant + suboptimal structures as shown in Figure 4.3.13.

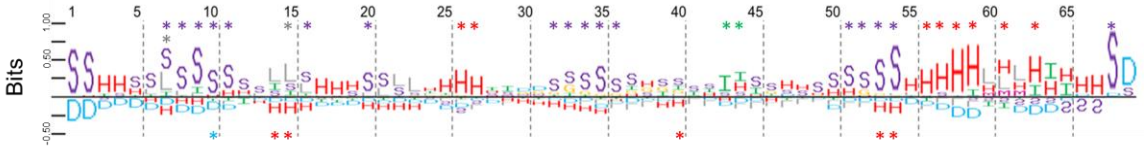


Figure 4.3.12. Position specific score matrix (PSSM) of SSE for dominant only structures of aptamers using the random sequence population as the background and α=1 in the additive smoothing function. A higher bit score either positive (i.e. stronger presence) or negative (i.e. stronger absence) for one or more SSE at a particular position corresponds to a greater height of the one-letter symbol for a given SSE. Base positions significant from Δ analysis are marked with an asterisk.

The dominant secondary structure PSSM Figure 4.3.12, shows close agreement overall in trends to the numerical differences shown in Figure 4.3.10. A notable departure in this trend, however, occurs near the ends of the sequence since there are fewer potential SSE possibilities at the 5' and 3' ends (e.g. a bulge cannot occur at base 1 or 69). Thus, any differences in SSE between the aptamer and the random sequence background become more heavily weighted. Another key difference is the inclusion of internal loops as significant near the 3' end. Hairpin loops also gained more weighed significance throughout, specifically at base positions 7, 14-15, and 60-62 as duplexes at base position 69.

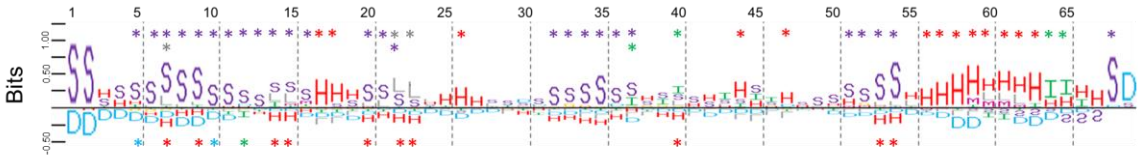


Figure 4.3.13. Position specific score matrix (PSSM) of SSE for dominant + suboptimal structures of aptamers using the random sequence population as the background and α=1 in the additive smoothing function. A higher bit score either positive (i.e. stronger presence) or negative (i.e. stronger absence) for one or more SSE at a particular position

corresponds to a greater height of the one-letter symbol for a given SSE. Base positions significant from Δ analysis are marked with an asterisk.

The dominant + suboptimal secondary structure PSSM in Figure 4.3.13 also shows overall agreement with its related analysis in Figure 4.3.11. Again, key differences observed are that SSE near and at the sequence end positions become more significant. Hairpin loops do not appear to gain as much significance as compared to the dominant structure PSSM, with the only exception being the hairpin loops occurring at base positions 22-23. Duplex formation at base position 69 was still considered significantly different from the background random sequence population.

A key advantage of the PSSM analysis is the ability to quantitatively score individual sequences to determine how closely they match both overall trends as well as one or more particular base positions. Scoring of the individual AuNR aptamer secondary structures, as shown in equation 4.3.15, allows the best match to the PSSM to be identified. The resulting scores of this separate analysis are shown in Table 4.3.2, with highest scores assigned to: 102.S1 (A7), 114.S1 (C9), and 219.S1 (D7), respectively ((XY) indicates the corresponding Figure 4.3.5 coordinate). Interestingly, aptamer SS$s selected from the A-rich random library do not score as high on average as the aptamer SS$s from the normal equibase random library. A separate PSSM that excludes all these A-rich aptamers (i.e. 4XX nomenclature) does not show distinct differences with the PSSMs shown in Figure 4.3.12 and Figure 4.3.13. This separate PSSM data is not shown here.

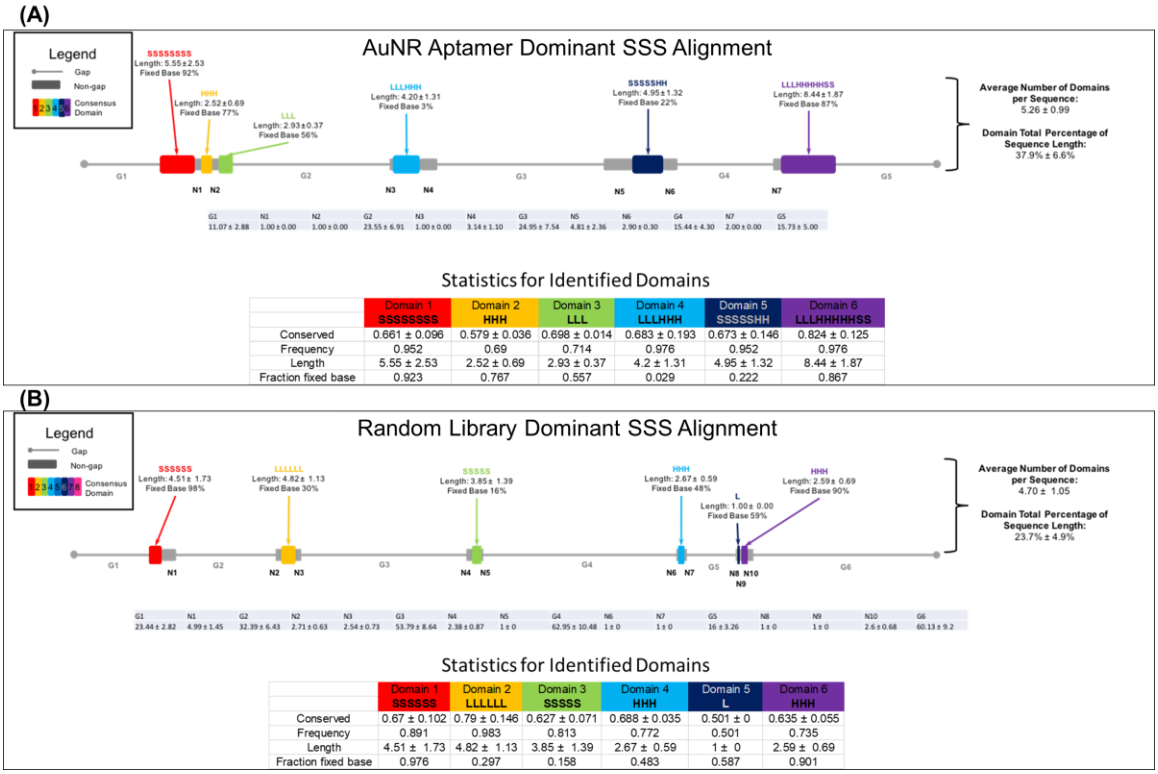
Table 4.3.2. List of AuNR aptamer SS$ and their corresponding PSSM score values τ_i for either dominant structures or dominant + suboptimal structures PSSM. Higher scores indicate a better overall match to the PSSM of SSE.

Aptamer SS$	Dominant and Suboptimal PSSM Score	Dominant PSSM Score
102.S1	15.20	14.94
114.S1	15.14	14.06
219.S1	14.98	14.02
111.S3	14.72	13.61
120.S1	13.91	13.71
204.S2	14.58	12.44
415.S2	13.52	12.42
412.S2	13.63	11.97
401.S1	11.98	12.16
110.S1	11.54	11.59
207.S1	12.72	10.40
204.S1	11.54	11.54
217.S2	11.78	11.13
116.S1	13.53	9.37
116.S2	13.43	9.04
212.S1	11.36	10.91
217.S1	10.55	11.06
214.S1	9.81	11.57
403.S1	12.08	9.22
112.S1	10.58	10.58
412.S1	9.98	11.15
105.S1	11.29	9.32
419.S1	9.80	9.97
117.S1	10.14	9.28
119.S1	8.90	10.46
417.S1	10.25	8.74
410.S1	9.81	8.73
409.S1	8.32	9.68
413.S1	8.51	9.33
201.S2	8.64	8.73
421.S1	8.92	8.44
420.S1	9.43	7.90
416.S1	7.76	9.49
111.S2	10.13	6.61
415.S1	7.27	9.13
121.S1	9.36	7.01
404.S1	8.89	7.32
115.S1	7.32	8.44
116.S3	9.03	6.47
101.S2	8.74	6.42
402.S1	8.72	5.80
108.S2	8.73	5.66
406.S1	7.88	6.44
414.S1	7.11	6.74
101.S1	7.07	6.26
118.S1	6.54	6.52
210.S1	7.02	5.65
108.S1	6.65	5.85
418.S1	6.71	5.06
111.S1	5.99	4.40
411.S1	5.61	4.49
403.S2	5.63	2.70
204.S3	5.03	2.10
405.S1	1.86	2.90
201.S1	2.98	1.24
112.S2	2.43	0.03
201.S3	-1.48	-1.41

4.3.6 Analysis of Aligned SSE through Multiple Secondary Structure Alignment (MSSA)

After using Mega7 with ClustalW to align SS$ across sequences, 1D representation of the global multiple sequence alignment results are generated as shown in Figure 4.3.14 for the dominant SS$s, and in Figure 4.3.15 for the dominant + suboptimal SS$s. With these 1D

representations of inserted gaps, domains in which a SSE predominates (present in at least 50% of the sequences), and non-gaps (where neither a gap nor a particular SSE predominates), it is more easy to display statistical data for the alignment.



	Domain 1 SSSSSSS	Domain 2 HHH	Domain 3 LLL	Domain 4 LLLHHH	Domain 5 SSSSSHH	Domain 6 LLLHHHHHSS
Conserved	0.661 ± 0.096	0.579 ± 0.036	0.698 ± 0.014	0.683 ± 0.193	0.673 ± 0.146	0.824 ± 0.125
Frequency	0.952	0.69	0.714	0.976	0.952	0.976
Length	5.55 ± 2.53	2.52 ± 0.69	2.93 ± 0.37	4.2 ± 1.31	4.95 ± 1.32	8.44 ± 1.87
Fraction fixed base	0.923	0.767	0.557	0.029	0.222	0.867

	Domain 1 SSSSSS	Domain 2 LLLLLL	Domain 3 SSSSS	Domain 4 HHH	Domain 5 L	Domain 6 HHH
Conserved	0.67 ± 0.102	0.79 ± 0.146	0.627 ± 0.071	0.688 ± 0.035	0.501 ± 0	0.635 ± 0.055
Frequency	0.891	0.983	0.813	0.772	0.501	0.735
Length	4.51 ± 1.73	4.82 ± 1.13	3.85 ± 1.39	2.67 ± 0.59	1 ± 0	2.59 ± 0.69
Fraction fixed base	0.976	0.297	0.158	0.483	0.587	0.901

Figure 4.3.14. 1D representation of Multiple Secondary Structure Alignment (MSSA) representation of all dominant predicted secondary structures for (a) AuNR aptamers and (b) random sequence population background.

In Figure 4.3.14(A), the MSSA for the dominant SS$s of the AuNR aptamers identifies six domains with the average number of domains 5.26±0.99 per sequence. On average, domains comprise 37.9% of a given SS$ with the remainder comprised of only non-gaps/non-domains. In comparison, in the random sequence population background in Figure 4.3.14(B) 6 domains are also identified in the MSSA, with an average number 4.70±1.05 domains. These domains comprise only 23.7% of a given random sequence.

Thus, while the same domains are identified in the random sequence population, the smaller average percentage of aligned random sequence involved in a domain may point to a large degree of disorder of the SS$ of the random sequences compared to the aptamer sequence, and shorter domains are identified. By examining the SS$ in each domain, it is evident that all domains involve bases from the fixed base segment – as high as 100% in domain 1 for aptamer in Figure 4.3.15. Additionally, on the 5' end of the alignment identical SSE are observed in SS$ in which Domain 1 is "SSSSSSSS" in the AuNR aptamer whereas in the random sequence population Domain 1 is "SSSSSS." This portion of the alignment is not surprising, as the distribution of SSE in Figure 4.3.9 indicates that single-stranded segments near the 5' end are predominant. There are greater distinctions near the 3' end, in which the SS$ in Domain 6 is of the aptamers that includes a hairpin loop, hairpin stem, and single-stranded segment. As previously shown in the PSSM, this predominance of a hairpin is expected; however, in the random sequence population Domain 6 is comprised of a single SSE, namely a hairpin stem. These domain difference between aptamers and the random sequence population appear even more distinctive considering the fact that ~90% the bases are from the fixed base segment. Even though for both the random sequence population and the aptamer MSSA, Domains 4-5 are comprised of a hairpin stem and loop, it is notable that for the random sequence population ~50-60% bases are from within the fixed base segment for the two domains, only ~3% and 22% of the bases are from the fixed base segment in the aptamers for Domain 4 and 5. This seems to indicate that while the fixed base segment on the 3' end has a tendency to form a hairpin stem and loop structure, as observed in ΔSSE and PSSM, the formation of a hairpin at this 3' end may be important to the overall secondary structure for many of the 42 aptamer SS$ as in

the alignment the hairpin is conserved even outside of the fixed base segment. In Domains 2-3, in the AuNR aptamer dominant MSSA another loop is observed; however this loop may be less distinctive since a hairpin loop also occurs in Domain 2 of the random sequence population. Domain 2 and Domain 3 in the random libraries do not involve a high percentage of bases from the fixed base segment as compared to Domains 2-3 in the aptamer. So in general, it may be favorable to form at least 2 hairpin structures in a given random sequence with base segments near the 5' ends tending to single-stranded.

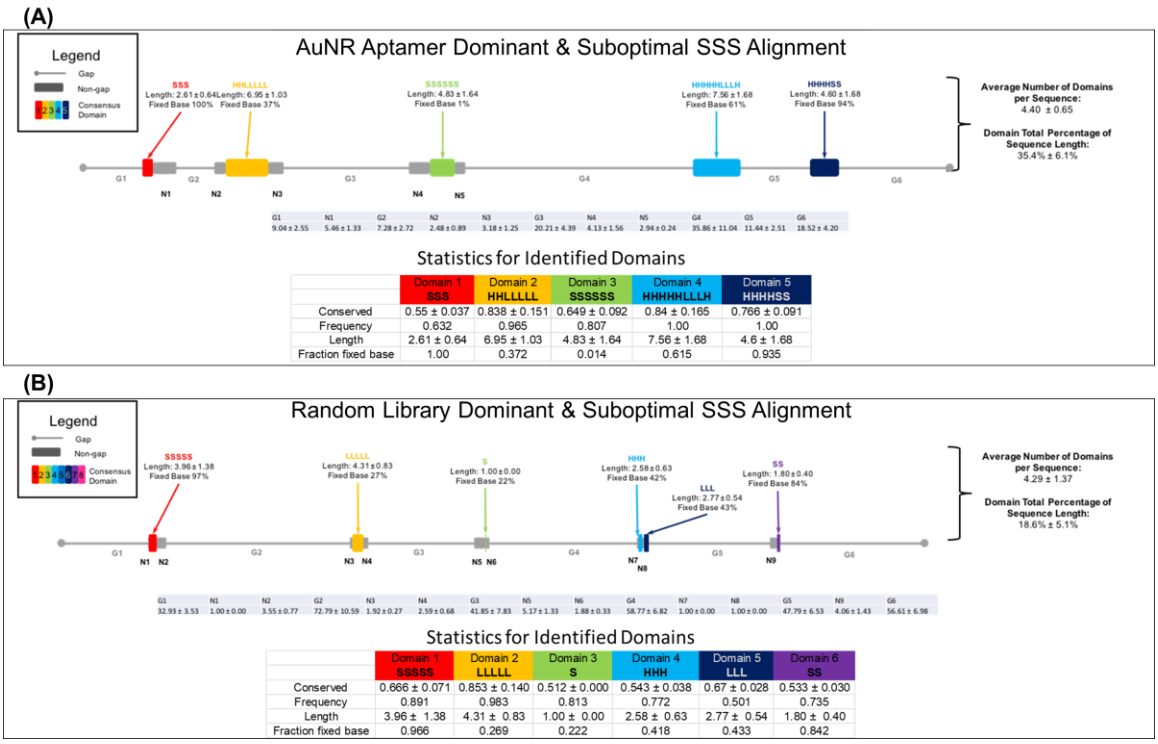


Figure 4.3.15. 1D representation of Multiple Secondary Structure Alignment (MSSA) representation of all dominant + suboptimal predicted secondary structures for (a) AuNR aptamers and (b) random sequence population background.

While in the dominant only MSSA for the aptamer, it was observed that an 8 “base long” single-stranded segment occurs in Domain 1, for the dominant + suboptimal MSSA of the

aptamer in Figure 4.3.15(A), a shorter (3 “base-long”) single-stranded segment occurs in Domain 1, involving only bases from the fixed base segment. The random sequence library alignment for dominant + suboptimal as shown in Figure 4.3.15(B) also continues to show prevalence of the single-stranded in Domain 1 in the 5' fixed base region. Domain 2 for both the random sequence population and the aptamer dominant + suboptimal structures, again show the formation of a hairpin for this domain. Domain 3, on the other hand, is a conserved single-stranded segment of six bases with ~98% of the bases occurring in the central segment for the aptamer, while Domain 3 only one “base” long in the random sequence population with ~78% of the bases occurring in the central segment. In the MSSA dominant + suboptimal structure of the aptamer, Domain 4-5 involve longer SS$ than even Domains 4-6 combined in the random sequence population, but the general trend of a hairpin stem to loop to stem to single-stranded can be observed to be spanning across all these domains, even though the other half of the hairpin stem (to complete the hairpin) is not conserved in the random sequence population. From these MSSA of dominant + suboptimal secondary structure, the general structure “skeleton” is likely “single-stranded to hairpin to single-stranded to hairpin to single-stranded,” or in SS$ condensed notation “SHLHSHLHS” in which the last hairpin stem and loop is typically located in the 3' fixed base segment. However, the central single-stranded segment of the aptamer in the central segment does seem more prevalent as a SSE as also suggested by both the ΔSSE and PSSM analysis.

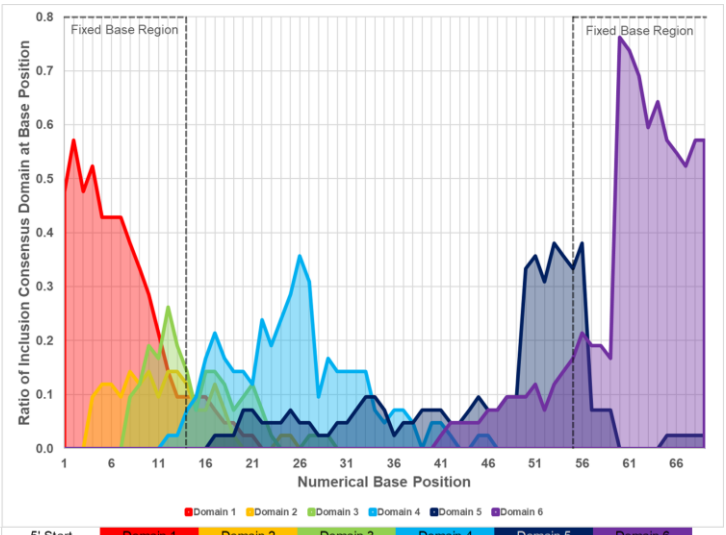
4.3.7 Base Position Dependent Distribution of Domains

MSSA of the SS$s thus far indicates that at least some structure is preferential for our templated sequence and that the aptamer prefers a single-stranded segment in the central

segment. To attempt to gain more information from the alignment, gaps were removed, while still retaining domain assignments to map their distribution within SS$s as a function of base position. From this distribution of domains, it is easier to identify segments in which structure formation is likely or favored for aptamer sequences. Other additional statistics on the identified domains from the alignment with the gaps removed is collected to potentially elucidate more information on secondary structure formation in the aptamer in comparison to the random sequence population. A distribution of the domains as a function of base position is shown in Figure 4.3.1616 for the dominant secondary structures. A number of key similarities and differences between the domain distribution in Figure 4.3.16(A) for aptamers and in Figure 4.3.16(B) for the random sequence population occurs. The first apparent similarity is Domain 1 with the distribution and SS$ looking nearly identical. The next apparent similarity is the location of Domain 6 as centralized around base position 66 and comprised of hairpins in the aptamers and random sequence population. The main obvious difference between the two distributions is the appearance of the peak for Domain 4 in the aptamer in Figure 4.3.16(A). While in the aptamer, Domains 2 and 3 have overlapping distributions and can be said to be very closely related as $d_{n,n+1}$ and d_{next} are identical values for Domain 2 (1.10±0.41) with Domain 3's d_{prior} also being very close in value (1.3±1.15). However, in the aptamer dominant structure domain distribution, Domain 4 does not appear to be completely related to Domain 3 with a d_{prior} of (9.34±8.4) while Domain 3 has a d_{next} of (7.37±8.31). These values indicate the conserved hairpin formed by Domain 4 within the random region is at least partially a conserved trait of aptamer dominant secondary structure. Lastly Domain 5 of the aptamer

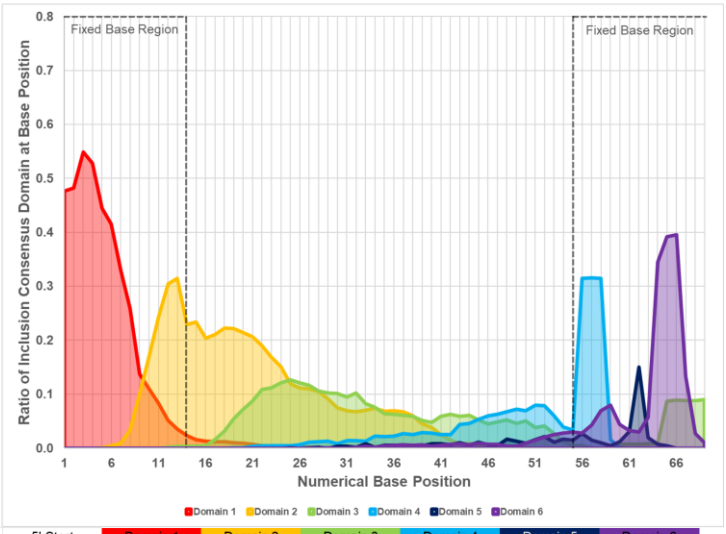
and Domain 4 of the random sequence population appear to be similarly conserved secondary structures comprised of a hairpin centralized around base position 56.

(A)



	5' Start	Domain 1 SSSSSSSS	Domain 2 HHH	Domain 3 LLL	Domain 4 LLLHHH	Domain 5 SSSSSHH	Domain 6 LLLHHHHHSS
d(n,n+1)	2.25 ± 3.39	2.11 ± 1.89	1.10 ± 0.41	7.52 ± 8.42	15.56 ± 9.78	8.54 ± 7.32	4.22 ± 6.31
d_prior			2.41 ± 2.49	1.30 ± 1.15	9.34 ± 8.40	15.25 ± 9.86	9.29 ± 7.97
d_next		5.30 ± 6.33	1.10 ± 0.41	7.37 ± 8.31	15.98 ± 9.77	8.30 ± 7.38	
domain loss		0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	1.07 ± 0.27	2.67 ± 1.53	2.00 ± 1.00

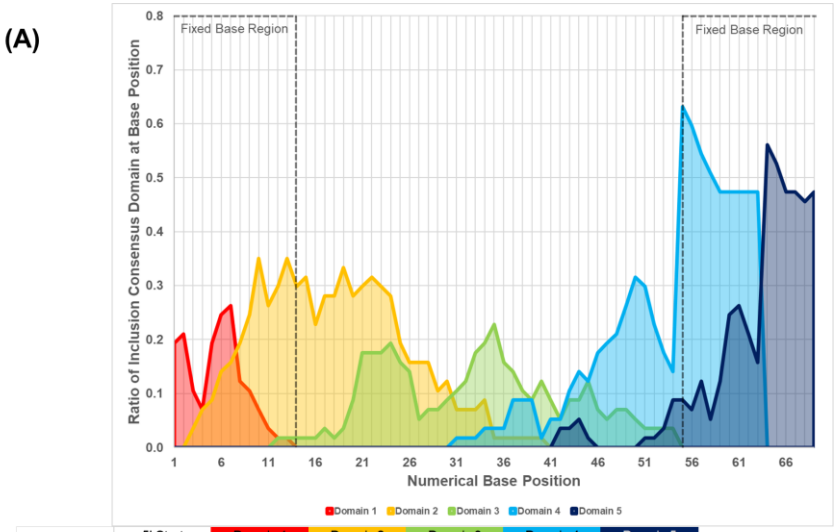
(B)



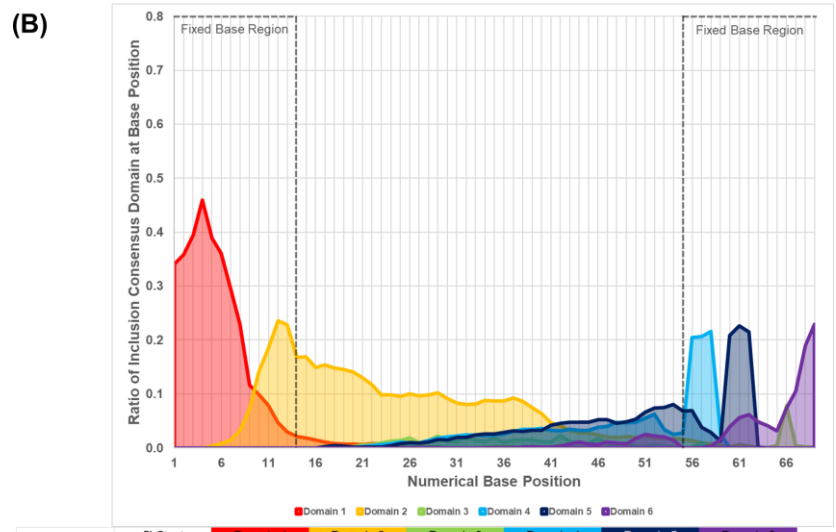
	5' Start	Domain 1 SSSSSS	Domain 2 LLLLLL	Domain 3 SSSSS	Domain 4 HHH	Domain 5 L	Domain 6 HHH
d_(n,n+1)	1.64± 2.8	12.02 ± 7.57	13.01 ± 8.65	14.11 ± 10.17	5.18 ± 3.96	1.13 ± 0.41	6.99 ± 7.18
d_prior			12.03± 7.55	13.12± 8.83	15.76± 10.87	7.15± 7.34	4.36± 6.25
d_next		12.19 ± 7.85	14.67± 9.89	15.17± 10.50	5.59± 3.50	5.34± 9.28	
domain loss		0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Figure 4.3.16. Domains determined from Multiple Secondary Structure Alignment (MSSA) are re-mapped without gaps as a function of base position for all dominant predicted secondary structures for (A) AuNR aptamers and (B) the random background. Tables are included in the figure which list average base length separation between various

neighboring domains as well as domain loss from the central portion of an identified domain.



	5' Start	Domain 1 SSS	Domain 2 HHLLLLL	Domain 3 SSSSSS	Domain 4 HHHHHLLLH	Domain 5 HHHHSS
d_(n,n+1)	3.25 ± 2.92	11.17 ± 6.07	11.09 ± 7.46	13.59 ± 10.78	2.65 ± 3.98	5.18 ± 6.58
d_prior			10.38 ± 6.38	11.13 ± 7.30	15.14 ± 11.82	2.65 ± 3.98
d_next		11.22 ± 5.99	13.2 ± 9.99	13.59 ± 10.78	2.65 ± 3.98	
domain loss		0.00 ± 0.00	1.22 ± 0.43	0.00 ± 0.00	1.53 ± 0.64	1.33 ± 0.58



	5' Start	Domain 1 SSSSS	Domain 2 LLLLL	Domain 3 S	Domain 4 HHH	Domain 5 LLL	Domain 6 SS
d_(n,n+1)	2.47± 3.74	14.14 ± 10.03	16.54 ± 8.53	9.55 ± 7.03	1.21 ± 0.54	10.07 ± 6.15	5.07 ± 6.9
d_prior			14.58± 10.49	16.7± 8.83	15.06± 10.17	5± 8.79	11.05± 7.7
d_next		14.33 ± 10.26	18.96± 9.71	10.86± 7.51	2.01± 3.76	12.18± 8.17	
domain loss		0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00

Figure 4.3.17. Domains determined from Multiple Secondary Structure Alignment (MSSA) are re-mapped without gaps as a function of base position for all dominant + suboptimal predicted secondary structures for (A) AuNR aptamers and (B) the random background. Tables are included in the figure which list average base length separation

between various neighboring domains as well as domain loss from the central portion of an identified domain.

In Figure 4.3.17, the case for dominant + suboptimal SS$s MSSA distribution of domains is examined. The first observable difference the distribution of the dominant + suboptimal MSSA domains for the aptamer in Figure 4.3.17(A) versus the random sequence population in Figure 4.3.17(B) is that Domain 1 in the aptamer is quite short as compared to the similar SS$ for the Domain 1 of the random sequence population and the much longer previous SS$ in Domain 1 of the aptamer dominant MSSA distribution in Figure 4.3.16(A). Additionally, the observed distribution of Domain 1 for the dominant + suboptimal SS$s of the aptamer appears to be bimodal as compared to all previous distributions for Domain 1. This "split" in the distribution of Domain 1 occurs as Domain 2 (a hairpin stem and loop) of the aptamer begins to appear, suggesting including suboptimal structures of the aptamer for consideration, hairpin formation near the 5' becomes more favoured and can extend to include bases from into the central segment. While the random sequence population involving dominant + suboptimal secondary structures for Domain 2 also includes for a hairpin, even accounting for the missing hairpin stem portion does not take away significance from the early formation of a hairpin being much more favoured in the 5' end of the aptamer and subsequently extending in greater proportion into the random region. For Domain 3 of the aptamer an almost bimodal behaviour of the distribution is also observed for its single-stranded segment identity, which centers itself into the very middle of the central segment near base position 35 or near base position 23, an observation which agrees with the PSSM for dominant + suboptimal cases. The Domain 4 hairpin of the aptamer appears to occur bimodally, either in the fixed base segment at base 55 or peaking around base position 50. While the random

sequence population also forms a hairpin with Domains 4 and 5, it is much less conserved as compared to the aptamer secondary structure domain distribution, suggesting that in the case of Domains 4 and 5 for the aptamer dominant and suboptimal secondary structures, the conserved hairpin near the fixed base region may be an important signature of the aptamer secondary structure, possibly acting as a structural "anchor" around which other secondary structure elements form. This analysis of the secondary structure distribution domains has also revealed that distributions of secondary structure domains within an aptamer appear to dramatically change with the inclusion of suboptimal structures that generally ignored as candidate structures.

4.4 Conclusions

We have developed a methodical tool and approach identifying patterns of motifs in secondary structure of aptamers, particularly once compared to a random sequence population. Through use of computational approaches, we have identified secondary structure families (SSF) and their difference in proportion to the random background. Base position dependent analysis of aptamer secondary structure has been used for differences in proportion of secondary structure elements (SSE) to identify key structural differences and provided the ability to score structure differences in aptamer candidates with position specific score matrixes (PSSM) compared to a background of random sequences. Additionally, multiple secondary structure alignment (MSSA) has employed used on aptamer secondary structure to compare to a random sequence background to identify differences in the distribution of secondary structure elements and demonstrate potential differences in an aptamers secondary structure distribution, especially if suboptimal secondary structures are considered. Data suggests conservation of a hairpin secondary

structure around or near the 3' fixed base segment, a centralized single-stranded segment in the variable base segment, and another hairpin forming around base position 17-19 may be structural signatures for identified AuNR aptamers.

CHAPTER 5. ADAPTING COMPELS SCREENING FOR DNA APTAMERS AGAINST A PROTEIN TARGET IN ITS OXIDIZED STATE

Natural and engineered antibodies are frequently used as binding agents for specific biological targets such as proteins. Many redox-sensitive proteins such as phosphatases, however, undergo transient changes that limit the practicality of antibodies as capture agents sensitive to the oxidation state of these proteins. Single-stranded oligonucleotides known as aptamers are gaining prominence as bio-inspired ligands and "capture agents" for a variety on non-nucleotide targets such as proteins. We hypothesized that the formation of intramolecular disulfide bonds between distal cysteine residues in MKP3 proteins during redox signaling create distinct tertiary protein conformations that can, in turn, regulate binding of oxidation state-specific DNA aptamer candidates. While the long-term goal (beyond this thesis work) will encompass identifying the oxidation site-specific DNA aptamer sensitive to for *transient* oxidation events, here we explore the ability to identify aptamer candidates for a phosphatase protein called MKP3 in its *irreversible* oxidation state using a non-evolutionary two-stage competition enhanced ligand selection approach.

.

5.1 Introduction

5.1.1 Key examples of historical protein-aptamer systems

Of the numerous biological targets chosen for various SELEX-based aptamer screening studies, thrombin and vascular endothelial growth factor (VEGF) represent two classic examples of protein targets for SELEX-based aptamer screening. The first aptamer sequence reported for thrombin known as the Bock aptamer[334] dates back to 1992 – just two years following the pioneering aptamer screening publications of Ellington and Szostak[53] and Tuerk and Gold[54]. Following a series of publications by Janjic and coworkers to identify RNA aptamer[114] and then modified RNA[115, 117] aptamers for VEGF, a chemically modified aptamer for VEGF called Pegaptanib is notable as the first FDA-approved aptamer[335] for therapeutic treatment of macular degeneration. Additional VEGF aptamer sequences have since been reported in the literature by other groups.[114, 118, 336] The popularity of the original Bock and Pegaptanib has fostered follow-up studies by other groups to understand the nature of binding between thrombin[308, 309, 337] or VEGF[338] binds to its aptamer sequence; however, the same attention to determining the structure-binding relationship is generally lacking for countless other aptamer systems. In particular, while the 3D structures of many proteins such as VEGF[339] are well-catalogued (see Figure 5.1.1(a) as an example), documented information regarding the conformation of the aptamer sequence itself – both alone and bound to its target – due to self-folding or self-hybridization of the oligonucleotide sequence is generally lacking. Most aptamer studies generally use self-folding programs such as UNAFOLD[18] to depict the aptamer as a 2D macromolecule, and then spatially place the non-nucleotide target such as VEGF (also typically depicted as a simple 2D

geometric shape, as illustrated in Figure 5.1.1(b), as binding to some segment of the aptamer sequence (without experimental verification of the actual binding motif in the aptamer sequence). In fact, in contrast to the thousands of screening studies to identify new aptamer sequences reported over the last two decades, a 2016 review article[340] reports that as few 16 structures of aptamer-target complexes have been published. Moreover, while structural information is more readily available for single-stranded RNA alone[315, 316] as well as bound to a protein[317, 318] (due in part to the physiological relevance of RNA as a single-stranded biomacromolecule) structural information on single-stranded DNA is far less comprehensive.

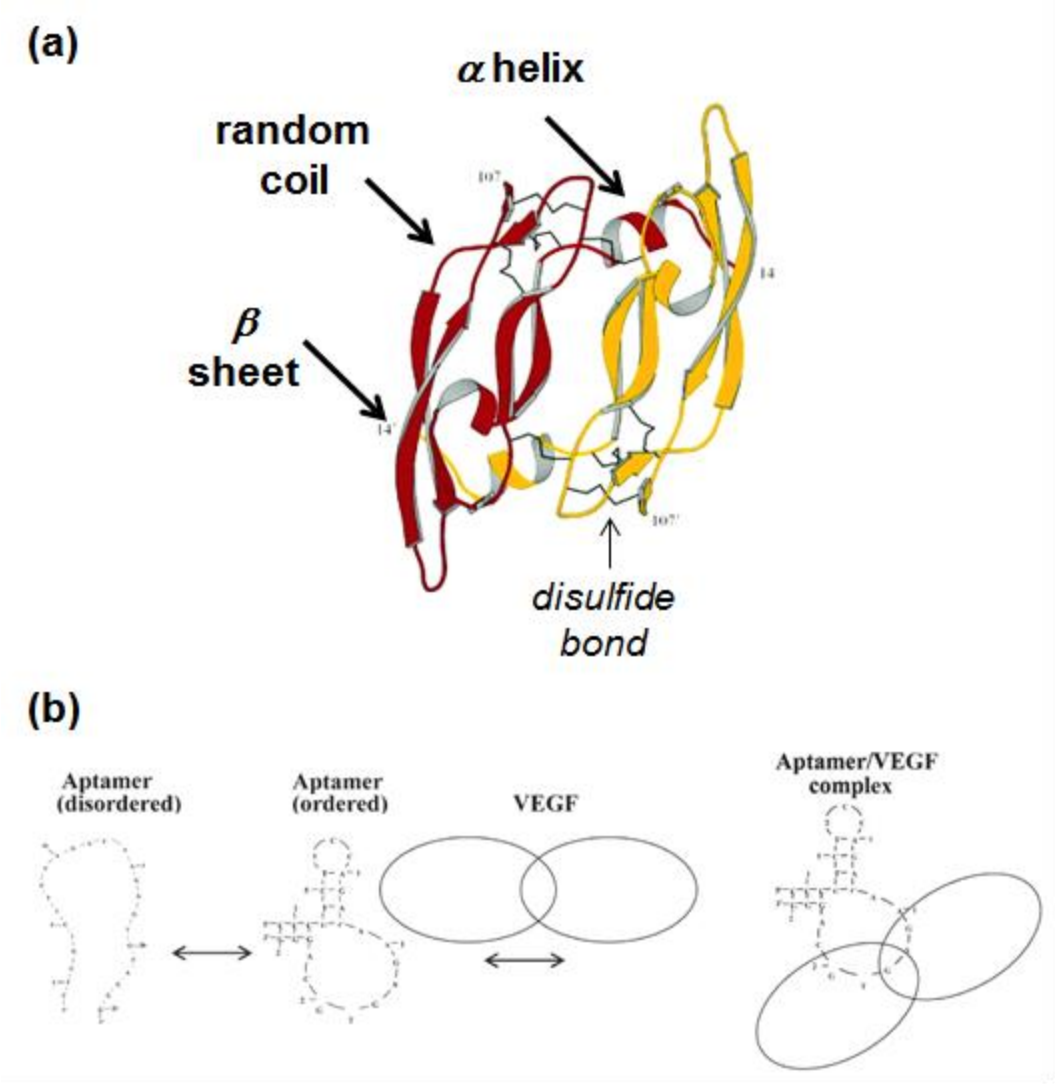


Figure 5.1.1. (a) Ribbon representation of vascular endothelial growth factor (VEGF) heterodimer (shown in red and yellow) with representative secondary structure elements of an individual b sheet, random coil and a helix labeled (Adapted from [339]). In contrast to the 3D structural detail of VEGF shown in (a), panel (b) shows a 2D schematic of VEGF bound to its self-hybridized aptamer reported by Gold and Janjić. NeXstar Pharmaceuticals, Inc[116]. Schematic in (b) is taken from [119]

5.1.2 Protein target that lack antibody options

While animal-derived antibodies can serve as alternative ligands for many proteins including thrombin and VEGF, other proteins can undergo dynamic conformational changes in physiologically-relevant conditions. The dynamic, often transient nature of

these conformational changes can thus inhibit the likelihood of identifying antibodies specific to that particular conformational state in the protein. The relatively small size as well as *in vitro* screening options for oligonucleotide-based ligands, on the other hand, could afford identification of aptamers sensitive and specific to these conformational changes in a given protein.

Intracellular oxidation is emerging as an important mechanism for the regulation of proteins called phosphatases across eukaryotic organisms. In mammalian cells, reactive oxygen species (ROS) are transiently produced during the course of normal receptor-initiated signal transduction. It has been recently reported that ROS production observed in some cell lines can cause inactivation of a family of phosphatases called protein tyrosine phosphatases (PTPs).[341] The PTP family contains a redox-sensitive thiol in a conserved active-site cysteine that reversibly inactivates the protein upon oxidation.[342] Mitogen-activated protein kinase or MKP3 is a PTP with an active site comprised of a cysteine residue (Cys 293) shown in Figure 5.1.2 (a) that forms an intermediate sulfenic acid (-SOH) illustrated in Figure 5.1.2(b). This intermediate is often referred to as a "backdoor" disulfide bond involving Cys 293 and one of multiple candidate cysteine residues far from the active site. This oxidation event induces a large, stable conformational change in MKP3 that has been structurally verified.[343]

While these oxidation-induced structural changes in PTPs such as MKP3 have been verified, current quantitative detection methods such as ELISA[344] require bulk lysates to measure actual phosphatase oxidation activity. Antibody options are limited since specific antibodies for this particular active site where these sulfenic acid intermediates reside are difficult to generate. Thus, while protein oxidation can be studied *outside* of the

cell (i.e. in bulk lysates), these restrictions preclude measuring these oxidation events *inside* cells. To identify candidate alternatives to antibodies that may serve as intracellular probes for protein oxidation events, the current work explores identifying and analyzing DNA aptamers against oxidized MKP3 ($MKP3_{ox}$).

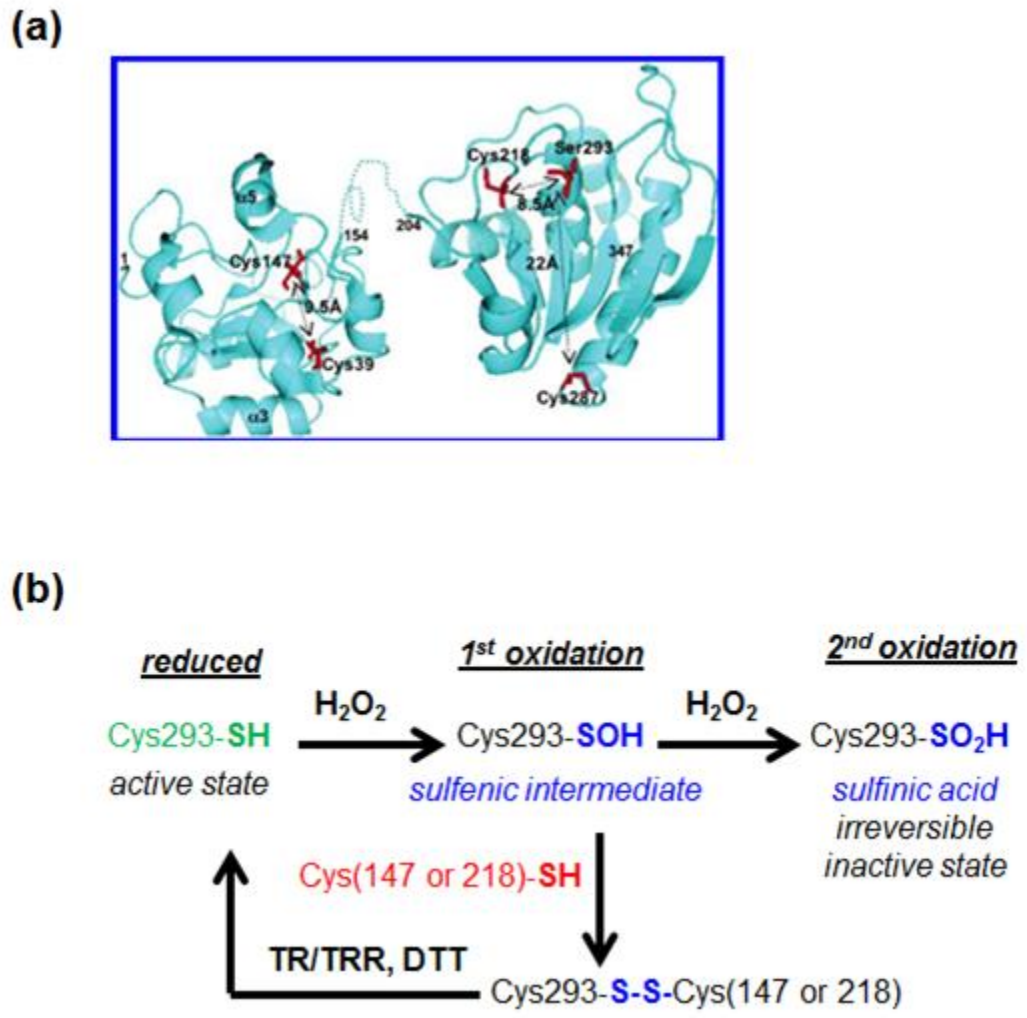


Figure 5.1.2. (a) Structures of MKP3 domains with key cysteine residues shown in red along with their approximate distances (Taken from [343]).(b) Schematic of possible oxidation and reduction pathways for MKP3 (Adapted from [343])

5.2 Materials and Methods

5.2.1 Materials for ssDNA Library, Oxidation, Reduction, and Selection

Random libraries consist of single-stranded DNA (ssDNA) templates 69 bases in length with a central 40 base random segment region between fixed base segments on the 5' and 3' ends to enable primer binding for PCR (5'-GGG ACA GGG CTA GC-[40N]-GAG GCA AAG CTT CCG -3'). Normal libraries composed of an approximate equivalent distribution base N (25% A, 25% C, 25% T, 25% G) were purchased from Integrated DNA Technologies (IDT, Coralville, IA). Random libraries were PAGE purified by the manufacturer while the phosphorylated reverse primer (5'-Phos-CGG AAG CTT TGC CTC-3') and forward primer (5'-GGG ACA GGG CTA GC-3') ordered from IDT were HPLC purified.

dNTP mix (10 mM), phenol:chloroform:isoamyl alcohol (25:24:1), ethidium bromide, TOPO TA Cloning Kit for Subcloning, One Shot TOP10 Chemically Competent *E. coli* and X-gal were purchased from Invitrogen (Grand Island, NY). GoTaq DNA polymerase and 5X colorless GoTaq reaction buffer were purchased from Promega (Madison, WI). H_2O_2, $CaCl_2$, HEPES, $MgCl_2$, $CaCl_2$ and KCl were purchased from BDH Chemicals (VWR Scientific, Radnor, PA). SOC medium, Ethanol and Tris EDTA pH 7.4 were purchased from Fisher Scientific (Pittsburgh, PA). DL-Dithiothreitol, ampicillin sodium salt, and dimethylformamide were purchased from Sigma Aldrich (St. Louis, MO). Agar (bacteriological), glycogen and LB broth (Luria-Bertani) were purchased from Amresco (Solon, OH). Lambda exonuclease enzyme and 10X lambda exonuclease reaction buffer

were purchased from New England Biolabs (Ipswich, MA). The MinElute PCR Purification Kit was purchased from Qiagen (Gaithersburg, MD). The pH 7.4 aptamer binding buffer (ABB) used for CompELS-based selection consists of 20 mM HEPES, 2 mM $MgCl_2$, 150 mM NaCl, 2 mM $CaCl_2$ and 2 mM KCl. All buffers and synthesis were prepared using 0.2 μm filtered water (18.2 MΩ-cm) from a Synergy UV-R ultrapure water purification system (MilliporeSigma, Merck KGaA, Darmstadt, Germany).

5.2.2 Library Preparation

Template sequences were amplified via polymerase chain reactions with either the random or A-rich template (0.17 pM), dNTPs (0.2 mM), forward primer (60 nM), reverse primers (60 nM), GoTaq polymerase (0.05 U/μL), and 1X supplied colorless GoTaq buffer adjusted to 3.5 mM $MgCl_2$. PCR was carried out on a G-Storm thermocycler with a 100 °C heated lid, with a 2 min hold at 95 °C followed by 25 PCR cycles (30 s denaturation at 95 °C; 30 s annealing at 47 °C; 30 s extension at 72 °C), and a final hold at 4 °C. An ethanol precipitation was performed on the resultant PCR product. Resuspended PCR product was digested with lambda exonuclease at 5 U/μg following the manufacturer's instructions to remove the phosphorylated hybridization partners. P/C/I extraction was performed on the digested PCR product and followed with another ethanol precipitation. Final ssDNA product was resuspended in aptamer binding buffer (ABB) and ssDNA concentration was adjusted to 2.0 μM and stored at 4 °C until used for screening.

Table 5.2.1. List of template DNA sequences with a central 40 base-long randomized segment (40N) flanked by fixed base segments (underlined) as well as forward primers and phosphorylated reverse primers used to prepare normal (i.e. equibase) random sequence libraries via solution PCR followed by lambda exonuclease digestion of phosphorylated hybridization partners to prepare the master, single-stranded DNA screening library. For cloning, non-phosphorylated reverse primer was employed.

DNA	Sequence
Template	5'-GGGACAGGGCTAGC-40N-GAGGCAAAGCTTCCG-3'
Reverse Primer	5'- CGGAAGCTTTGCCTC-3'
Forward Primer	5'-GGGACAGGGCTAGC-3'

5.2.3 Materials for Coupling MKP3

Rabbit anti-His 6x polystyrene beads were purchased from Spherotech (Lake Forest IL USA). Tris base was purchase from Promega (Madison WI USA). NaCl and $MgCl_2$ was purchased from VWR (Radnor PA USA). EDTA was purchased from Boston Bioproducts (Ashland MA USA). NP-40 was purchased from US Biological (Salem MA USA). Glycerol and leupeptin hydrochloride was purchased from Sigma-Aldrich (St. Louis MO USA). PBS was purchased from Boston Bioproducts (Ashland, MA USA). Hepes was purchased from Mediatech (Manassas VA USA) KCl was purchased from EMD Millipore (Billerica MA USA). pepstatin A was purchased from EMD Biosciences (San Diego CA USA). MKP3 protein was purchased from Abcam (Cambridge UK).

5.2.4 Coupling of the MKP3

MKP3 was linked with antibody-coated polystyrene beads, ensuring unidirectional positioning of the protein on the beads via an N-terminal His tag. A sample of 100 μL rabbit anti-His 6x polystyrene beads was first equilibrated with 1 mL pH 7.4 equilibration buffer containing 25 mM Tris, 150 mM NaCl, 1 mM EDTA, 1% NP-40, and 5% glycerol for 30 min under slow rotation at 4°C. Following this, the beads were centrifuged for 2 min at 200 x g, 4°C, rinsed twice in sterile PBS and resuspended in 250 μL binding buffer, pH 7.6, containing 25 mM Hepes, 0.1 mM EDTA, 100 mM KCl, 12.5 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 10 μg/mL pepstatin A and 5 μg/mL leupeptin hydrochloride. Following resuspension, 10 μg of full length recombinant human MKP3 protein was added to the bead sample and the sample was incubated at 4°C overnight under slow rotation. The following day, the beads were centrifuged as above, wash three times in sterile PBS, and resuspended in 100 μL sterile PBS with pepstatin and leupeptin at the concentration above. The beads were stored in -20°C and washed with sterile PBS prior to use. (Thanks to Dr. Linda Kippner for performing MKP3 coupling.)

5.2.5 Preparing Counter Selection Target

100 μL of rabbit anti-His 6x polystyrene beads at 0.1% w/v were washed 3X in ABB and resuspended in 2 μM dNTPs in ABB and incubated for 30 min on rotomixer immediately before screening.

5.2.6 Reduction of MKP3

150 μL of MKP3 coupled polystyrene beads at ~7*10^6 particles per mL by hemocytometry were aliquoted into a PCR tube and centrifuged for 3 min at 9.1 kG and the supernatant was removed. (wash step) Beads were resuspended in 100 μL of 50 mM DTT in ABB with and incubated on ice for 30 min. After 30 min incubation with DTT, beads were washed 3X in 100 μL ABB. When the negative counter target had finished its 30 min incubation with the 2 μM dNTPs in ABB, a separate aliquot of 100 μL of 2 μM dNTPs in ABB was incubated to the reduced MKP3 beads for 30 min. After 30 min, reduced MKP3 beads were washed immediately before screening.

5.2.7 Oxidation of MKP3 Target

150 μL of MKP3 coupled polystyrene beads at ~7*10^6 particles per mL by hemocytometry were aliquoted into a PCR tube and centrifuged for 3 min at 9.1 kG and the supernatant was removed. (wash step) Beads were resuspended in 100 μL of ABB with 1mM H_2O_2 in ABB and incubated for 30 min. After 30 min incubation with H_2O_2, beads were washed 3X in 100 μL ABB. When the reduced MKP3 beads had finished its 30 min incubation with the 2 μM dNTPs in ABB, a separate aliquot of 100 μL of 2 μM dNTPs in ABB was incubated to the reduced MKP3 beads for 30 min. After 30 min, oxidized MKP3 beads were washed immediately before screening.

5.2.8 Aptamer Screening with MKP3 targets

ssDNA library was separated into 10 aliquots of 100 μL in PCR tubes and denatured in the thermocycler with heated lid (100 °C), 90 °C for 10 minutes, 4 °C for 15 min, and 24 °C for 5 min. A new ssDNA library aliquot was added to the counter selection target after every 30 min incubation and wash step of the counter target. Removed supernatant from

the counter target was added to the reduced MKP3 ($MKP3_{red}$) beads after every 30 min incubation and wash step of the $MKP3_{red}$ beads. Removed supernatant from the $MKP3_{red}$ beads was added to the oxidized MKP3 ($MKP3_{ox}$) beads after every 30 min incubation and wash step of the $MKP3_{ox}$ beads. The process of sequential removal and addition of supernatant is shown in Figure 5.2.1. This process was repeated until all 10 aliquots had been incubated with the oxidized target. Following the last of the 10 target-library incubations and wash, 200 μL of ABB was added to the reduced and oxidized bead suspensions followed by centrifugation and supernatant removal. These wash steps with ABB were repeated for a total of six washes. Following the six washes, nanopure water (64.7 μL), freshly prepared 20 mM $MgCl_2$ (10 μL), 5X GoTaq buffer (20 μL), dNTPs (2 μL, 10 mM), non-phosphoylated reverse primer (1.2 μL, 5 μM), forwared primer (1.2 μL, 5 μM), and GoTaq (1 μL, 5U/μL) were added to each the $MKP3_{red}$ beads and the $MKP3_{ox}$ beads. PCR cycling was carried out as detailed previously for the ssDNA library preparation. After cycling, dsDNA PCR product was stored at 4 °C until purification for vector insertion.

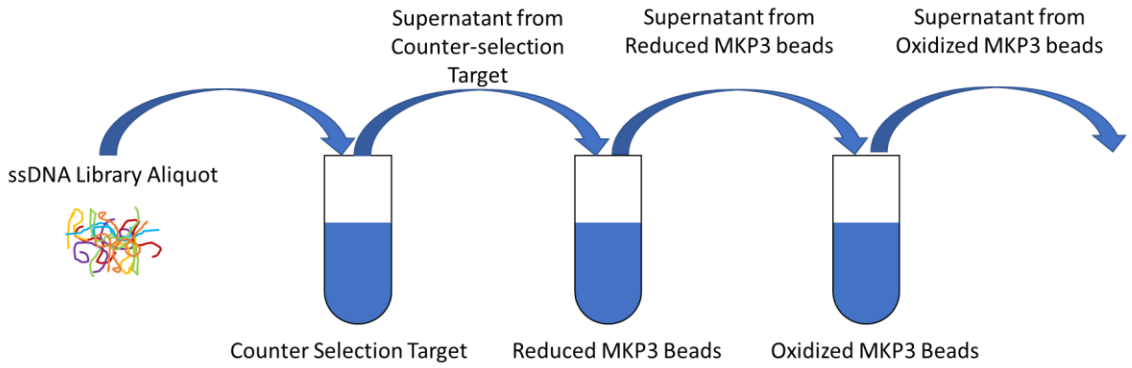


Figure 5.2.1. Process for serial additions of supernatant containing ssDNA library aliquot from the prior selection target to the next selection target

5.2.9 Vector Insertion and Sequencing

A Quiagen (Gaithersburg, MD) PCR purification kit was used and purified product was resuspended in 10 uL of 10 mM Tris-HCl pH 8.5. The 10 uL of purified PCR product was placed on ice and salt solution (2.5 uL, 1.2M NaCl, 10 mM MgCl2) and 2.5 uL TOPO vector from TOPO TA Cloning Kit (Invitrogen, Life Technologies, Grand Island, NY) was added. This ligation mixture was placed on a thermomixer at room temperature for 15 min at 500 rpm. 5uL of the ligation mixture was added to the TOP cells, gently mixed, and put on ice for 2 h. TOP cells and ligation mix are then heat shocked at 42°C for 30 seconds. 250 uL of SOC medium was added to each vial. Resulting cell suspensions were incubated on a shaker table incubator incubator for 1 hat 37 °C and 250 rpm. Transformed bacteria are then plated on LB-agar medium supplemented with ampicillin and X-gal, followed by overnight growth in a 37 °C incubator. Three resulting plates from each the $MKP3_{red}$ or the oxidized $MKP3_{ox}$ were sent to GENEWIZ Inc. for sequencing analysis (South Plainfield, NJ).

5.2.10 Checking Recovered Sequences for Errors

Identified sequences returned from GENEWIZ Inc. were checked for errors in their fixed base segments, random region length (e.g. missing a base in the 40 base-long segment), or unresolvable non-determinate bases. Any sequence possessing one of these errors was removed from consideration. Any non-erroneous complementary sequences to the template provided as a recovered sequence reverse transcribed (analytically) to yield the original template.

5.2.11 Primary Structure Analysis

The primary sequences of identified aptamer candidates are analyzed and aligned using T-Coffee (freely accessed 6/24/17) multiple sequence alignment of the 40 base central randomized regions (bases 15-54) of the sequences to identify and compare position-dependent bases among all aptamers. MEME 4.9.1 (freely accessed in 6/5/2017 at http://ebi.edu.au/ftp/software/MEME/) was then used to identify position-independent sequence segment motifs that occur anywhere within the central 40 base variable segment of the aptamer sequences.[172, 175]

5.3 Two Stage CompELS for Complex Protein Targets

Aptamer selection for both $MKP3_{red}$ and $MKP3_{ox}$ was performed in two separate, but simultaneously occurring stages. Before introduction to either protein target , a counter selection target consisting of the immobilization substrate was used to inhibit potential selection of an aptamer for the target substrate consisting of the rabbit anti-His 6x

polystyrene beads used for coupling the MKP3. An overview of the selection process used is shown in Figure 5.3.1

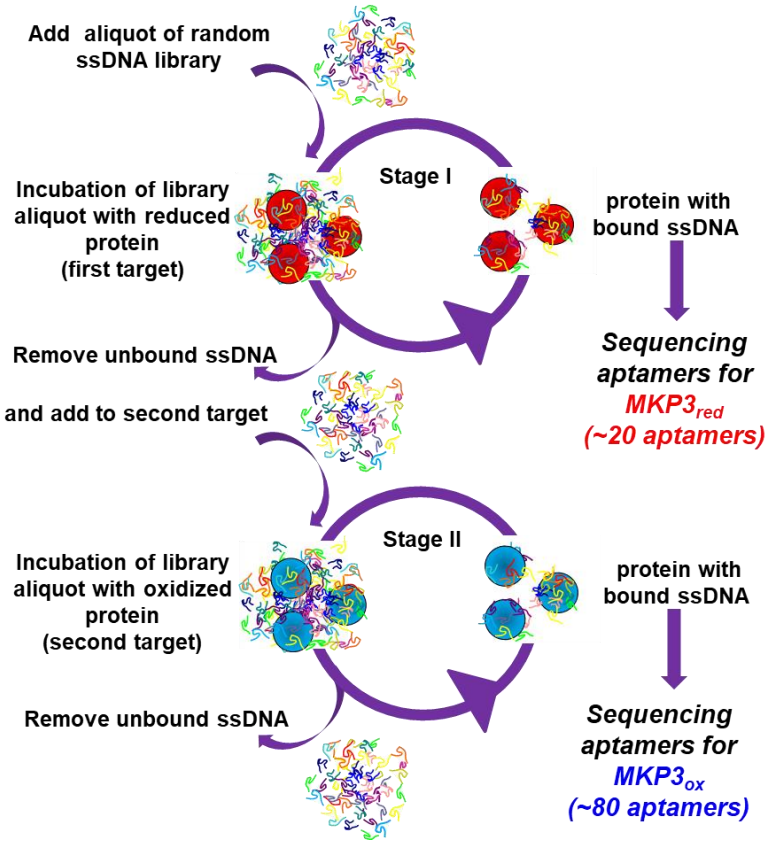


Figure 5.3.1. Overview of Two-Stage CompELS Aptamer Screening Approach against Oxidized MKP3. In stage I, selection against the reduced form of MKP3 is carried out. In stage II, remaining unbound ssDNA from stage I is used for selection against the oxidized form of MKP3.

The key difference to this work from the AuNR target is the dynamic yet closely related reduced and oxidized states of the biomacromolecule targets compare to a metallic particle substrate. With this two-stage process, aptamers specific to the final target of the $MKP3_{ox}$ could have less affinity to the $MKP3_{red}$, due to having been first incubated with the $MKP3_{red}$ prior to $MKP3_{ox.}$ For multi-target screening CompELS compared to SELEX, has the benefit of both being extremely time efficient for screening itself, but also has the potential to

identify aptamers that have the ability to distinguish between conformationally related targets as well as identification of aptamers for each target in one screening session without intermittent amplification steps. As indicated in Figure 5.3.1, in this selection over 100 aptamer sequences were identified for MKP3 with 23 sequences for the $MKP3_{red}$ and 100 for the $MKP3_{ox.}$ Sequences identified for $MKP3_{red}$ are given in Table 5.4.2 and sequences identified for MKP3ox are given in

Table **5.4.1**.

5.4 Resulting Sequences from CompELS of $MKP3_{ox}$ and $MKP3_{red}$

Table 5.4.1. List of nomenclatures and central 40 base-long segments of CompELS-identified aptamer candidates selected against oxidized MKP3 ($MKP3_{ox}$) The full-length 69 base-long sequences include two fixed base segments and a center variable region (40N) as follows: 5'-GGGACAGGGCTAGC(40N)GAGGCAAAGCTTCCG-3'.

Nomenclature and Sequence of Aptamer Candidates for $MKP3_{ox}$	
+3-01	**AAATGGCTAAGATATTTTGTTTAAATCGTCATCTATGGAC**
+3-02	**TATCCTGGTACCGACTCACGACACTGATAATAAACTGTAA**
+3-03	**CGAGTGGATCCACTTCTTTTTGGTTAGCTTAGTTAGTTTA**
+3-04	**CTTGTTTCCTTAAAGGTACAAGAAGACTGCACCCCAGGTT**
+3-06	**TGGACTGATATAGTGGAAGGAACCCGTTCTTGGAAAACCT**
+3-07	**AAATGGCTAAGATATTTTGTTTAAATCGTCATCTATGGAC**
+3-08	**AAGAGACGTTTAACCTTGGATTAATAATGAATACAAATTA**
+3-09	**TCGTATACCGATTCCGAACTCATATCTCATACTAGCAAAC**
+3-10	**ATTTGGAGTGGAGGTTATACTTGGTGAGATGAATAAGTTA**
+3-11	**GACAAGAAGATCATCTCCAGGGCAAAGTCAGCCTCNAAAT**
+3-12	**TCTGGCTCACCTTTAGAAGTTAACTTCTATCGTGTAGTTA**
+3-13	**CGAACTATGCTGAGGCCCGGCAGATAATGATTACACCAAC**

Table 5.4.1 continued

+3-14	AAAAATAAATGCCTTCAACAAGGATACTGGTTCACGCTTT
+3-15	GAGATTTTTAAGGTCAACTTTTAATGGCATTTTGAGAAAA
+3-16	GTTTAAACTTATAACACTGAAGTGAACCCGGGGTGGGATT
+3-17	TTTTAAGATTAACCGACTCTGGCTTTTGTTTCGACCGTTC
+3-18	GCCACAAGTAATATCCTGCCGAATTATTTAAATTCACAGA
+3-19A	AGAAGTTTGATTAAAGAGTCGAATGAGGGACTGGTAATTA
+3-19B	AAATGTAAATTCGACACCCGTGTCGGAGGGAGGGGTATCA
+3-20A	CGCAAGGTATTTTCGAGTGGTAATTACCTATTTTAATCGC
+3-20B	ATCTCATGCAAAGGCATAGGCTCAAATGGGATAGTCGAAA
+3-21	GGTAAACCTTGAGATTCGTTCTTTTGGGTAATCAGCTCAT
+3-22	AGCTGGAACTTTATTGACATCCGACAAATGAGAATTTCTA
+3-23	TTGCGATGATCAGGGTATGAGGCAGGAGAAATAAAGAAAC
+3-24	CGAACTATGCTGAGGCCCGGCAGATAATGATTACACCAAC
+3-25A	TCAAACTGTAGCTCCTTCTGAGTAAATTCCTTGCTAGTTC
+3-25B	AATGGTACGCCATTTATGGAGGCAATTATACAACGAAATT
+3-26	CGAACTATGCTGAGGCCCGGCAGATAATGATTACACCAAC
+3-27	GTTGTCATTTGAGAGGGTAAATCGATGATGGGTAATTTCT
+3-28	GGGCCCAAAATGTAAGATCTTGGACCAACTTTTAAACAAA
+3-29A	CTCGTATAAGAACACATAAGGTTGTATCCTACCCTAAGTT
+3-29B	TACTACTCATGGTCTTGCGGCTGTGGTGGGGGCTGTTTAT
+3-30	CCGTGGGCGATGTTTAAGATGAAGTGGGATATTTATACTT
+3-31	GGGATTGGTATCTGTGGAGAAAAAACGCTGCTACAATAGA
+3-32A	GAATGGTAGTAAGGGGGACATGTAGGTCGGGGGCCGAATA
+3-32B	GGCATGACACGAGACGGCTATCAGGGGGAGCGGTGGCACC
+3-33	TCTGGCTCACCTTTAGAAGTTAACTTCTATCGTGTAGTTA
+3-34	CTTTATGTACGTGAATTATTTATATTGGGACCAGGGTCAA
+3-37	GGTAAACCTTGAGATTCCTTCTTTTGGGTAATCAGCTCAT
+3-38	CCAAAATTTTGATCTACGCGNTTCACAATTAGAATCTCTT
+3-41	TAAACTCCGCTATTAAAGAGATCGAGTAGGAATTCGTCTT

Table 5.4.1 continued

+3-42	GTTTGAGACTTAGTGATCCCCTTGCGTATTATGTACTATT
+3-43	CATGTGCTAATCATATGGTGTAGTATTAGGACCCAACGTC
+3-45	ATCTAGTCTATGTTAATCTAGGATGAAGATGTTGTTAAAC
+3-47	CACGATCGTGAAGACGACTCGATAGGCCAAAAGAGAAGTA
+3-48	TATCCTGGTACCGACTCACGACACTGATAATAAACTGTAA
+2-02	CGAAATCCTAAGTTAATTTTAGAGGGAACGAGGCCCCATC
+2-05	AGACAGAGCACTTAGTTTTCTATTAATAGCGCAATCTCTC
+2-06	TAGACGACACGATAAAGGGGTTCGCCGTAAGGCTGGAACA
+2-08	GGGGCAAGTGGATATGATGAAAGGTGGATCTGGGGGGATT
+2-09	CTCAGTGAATACACCGGTGCTGGAGTACAAGCGCGTTAAA
+2-10	GGGGCAAGTGGATATGATGAAAGGTGGATCTGGGGGGATT
+2-11A	TCGCTTTGTGAAGGGTGCCAGAGTGGGGGCGGGTTGATGA
+2-11B	AATACTTCAACCTATCTAACGTTTTCCTGCAAAAACCATA
+2-11C	AAATCATAAGAAAGGTTTTGAGCCTGCGTATGTTCATCAT
+2-12	CAGGAAGTACCCGAAGTATTATAATTTTGAACTACATTCA
+2-15	GAGCAAGGAATTCGTGAGATAACGTCTCATCATATTAAAC
+2-17	AGACAGAGCACTTAGTTTTCTATTAATAGCGCAATCTCTC
+2-18	GCCTTTACGATACGGTCCTGGTAGACTCTTATATGCGATC
+2-20A	TAAACGTTCTTAGTTTACGTACCTGTACAAGAGCGATTCC
+2-20B	CGAGCATCCATTGTGTCTTAATGAATAGTGCCTTATCTTA
+2-25	TCACGTAAATGGGTATTAGGACCGCATCATCACGGTAAAA
+2-31	TGGATTTTGAGTGGAGGGGATCATTTATTGGTGTCGTGAC
+2-35	TCACGTAAATGGGTATTAGGACCGCATCATCACGGTAAAA
+2-36	ACTAGCGCGGTTGACCGAAAACGATTGTAGATTGTGAAAT
+1-01	CGTATTAAGATTAATGGTATACTCCAGCCATTAATATGGA
+1-03	AATTTGTGCGACGCTTTCCGCCATGTTAGGATTGTCTTGA
+1-04A	TTCATTTAATAATGTAGAATCAGTAAGATTGACGATGCTT
+1-04B	TATGGCTCTCTACTAACACTGGCAACTCAAGGCGATGCAC
+1-06A	ACAACCTTTCCGTTATGGGTACCCCAAAGTCACTCTTTAT

+1-06B	CTTAGACTGTTTTCTAGTTAGAATGTCTTGCGAGTGCTTA
+1-10	AAGAATGATCAAAATTAGCAATAACGTTTCTTTATCAAGT
+1-13	ATGTTGGGGTTGTTAGTGTAGGGGAAATTCCTATTGTGAC
+1-14A	GTAATAAAGGTAGCTGGCAAATTTGGAATATACTTTCATC
+1-14B	TTAAAAAAAACGTATTTCCGAGTAGCAGTTCACGAGTAGT
+1-15	TACGTTTTTCTTATTACCCATATAATTGGTGTAATGTAAT
+1-17	ATAGTTCCCTATAAATCCCGGGTCCTGGGTGTGTTACTCA
+1-21	GATGTCTCCGGCCAGGAGTGTACTAGTAGATAATGGTGAA
+1-22	AAAGGATTTAGTTCTTAAATTAGCATCATTCAAGATCCAA
+1-25	AAGAAAATGCTGATCTTGGAAAACTTTAATCTTTAAGTCT
+1-26	ACAGTCTGGCTCTGGGTTAGATGCTAGACCATGATATCTC
+1-29	CTTTTGTGAAGGCAGTATTCCAATTTTATACGAGGCTTAA
+1-30	AACGTTAGTTTAAGGTTAGTCTCCTTCAGTTCACCAAATC
+1-32A	GTAGTGTAACGAGATATAGGGGAGGGGCGGGAAAGGTATA
+1-32B	ACTTTAGATTGAGAACAATACAGACGCCAATTGACAGGCT
+1-33	CTTTTTTGTTTTCAACTGAATCAACGAATACTTCTGAGGC
+1-35	ATAATAAGGCTGATTTTGACTTCAACAGTTGTCTTGTATT
+1-36	AGATTAAAACTTTAACAGTCTGAAATAATAACGATCAAAA
+1-37	GCTGTTACCGAGCATAAAGTCAACTTCGGTATATAGTGAT
+1-38A	CTGGGCGCTTTAAAGATTTGTGATTCCTGTGCTCATAAAA
+1-38B	AATGTGGATAGAGTTATAAGCCTCGTTCGGGTCATATTAT
+1-40A	TCTGTTCGCGTCAAGATTAAATTGTATTCGATTGGAGTAC
+1-40B	TCAACAGTATCTTATTTAGACTGTAGGTTTAACCTTATCC
+1-41A	TTCATTTAATAATGTAGAATCAGTAAGATTGACGATGCTT
+1-41B	TATGGCTCTCTACTAACACTGGCAACTCAAGGCGATGCAC
+1-49	AAGAAATGCTTACCTGCACGCACCTTCATGTATATAGAAT
+1-51	CGGCCATAGATATTCAGGCTAGGCTCATCTACTTCGTTAC
+1-52	CAATCAGAGAACGCCATTTACAGAACTTTAATAAAACAAC
+1-55	AAGAATGATCAAAATTAGCAATAACGTTTCTTTATCAAGT

+1-56	CGGCCATAGATATTCAGGCTAGGCTCATCTACTTCGTTAC

Table 5.4.2. List of nomenclatures and central 40 base-long segments of CompELS-identified aptamer candidates selected against reduced MKP3 ($MKP3_{red}$) The full-length 69 base-long sequences include two fixed base segments and a center variable region (40N) as follows: 5'-GGGACAGGGCTAGC(40N)GAGGCAAAGCTTCCG-3'.

Nomenclature and Sequence of Aptamer Candidates for $MKP3_{red}$	
-1-01	CCCTAGGTGACACTCCAACCTCAGAATTAATGTGTTTAAT
-1-04	GGACTTTGGGAATATCATGCGGGGTTTGGATAGATGATAA
-1-05	ACTCTAAAAATAGATTGTTGTTAATTTAACCTATAGGCCC
-1-06	ACTCTAAAAATAGATTGTTGTTAATTTAACCTATAGGCCC
-1-09	GGACTTTGGGAATATCATGCGGGGTTTGGATAGATGATAA
-1-17	ACTCTAAAAATAGATTGTTGTTAATTTAACCTATAGGCCC
-1-18	GTTTGTTATGACCCGAATGTTGGGCAGAACGAGTTGGGAT
-1-19	CACTCGGTCGGTTACTGTCGTGCTCTCATCAACGTATGGC
-1-21	CCCCTTCAATTTACTTCACTCCTAATTATGTCGTACCGTA
-1-22	AATGCCTTACAACGACAGGACTCAGACACGTCTTCGGGTC
-1-23	TGCAACAAGAAACGCGCAAGGATAGTTAGGATACTGCAAC
-1-25	GGACTTTGGGAATATCATGCGGGGTTTGGATAGATGATAA
-1-32	AATGCCTTACAACGACAGGACTCAGACACGTCTTCGGGTC
-1-43	TGCAACAAGAAACGCGCAAGGATAGTTAGGATACTGCAAC
-2-02	TAGCAGTGATAATGCTCAGCTCCATTAATTAGGAGAAGTT
-3-05	AAAGGCCGGAGTATGATGGATTCAAACCGTTTCGCTCCTA
-3-12	TTTAACGTAGTAAACTTTTCAAGTGTAAATCTCACAATAT
-3-13	TTGTAGTACTTTTACGTTTAGATGAGGTTGGGCACCATAA
-3-14	TGGGTAAATCGGTGGAGTAGATAAGTTATCGACGGGTCAT
-3-19	GTATCACTGTTGAGGGATTTATGGCGGACCCCAGAGGCCC
-3-24	ATCAATTTCAATCATATACACATAATGGTTTAAGTCTCAA
-3-29	GCCAGGTAAATAACGCACAAGGTCCACTATTCTCCTATTT

-3-30	ACGTTAGTGATTTTATCCCTAAGTTTATCTATTAGGTCAA

5.5 Primary Structure Analysis

5.5.1 MEME Analysis of MKP3 Sequences

These identified sequences are analysed for position independent primary structure motifs using MEME, in the same manner as described in Chapter 3.

In Figure 5.5.1, the 10 statistically significant motifs identified are shown. Since the aptamer sequences are color-coded it is easy to distinguish that several motifs consist of identical sequences though nomenclatures are distinct. It should be noted that MEME analysis was also performed for the case with no identical sequences included several statistical significant consensus motifs were also identified (not shown), so to more broadly examine representative sequences recovered from the screening, no sequences were removed from the list of aptamers for consideration in primary structure comparison. A similarity between motifs reported from MEME is that they seem to be extremely conserved across the random segment with only 2 of the 10 reported motifs having less than the full 40 random base segment being involved in the motif. Interestingly, the second identified motif, the guanine base is highly conserved with 24 of the 40 bases being possibly associated with guanine. 5 out the 10 identified motifs involve motifs made from both MKP3ox and MKP3red aptamer which indicates that there is likely a high degree of similarity in primary structure for aptamers for both targets. To further investigate correlations between identified aptamers, an alignment was performed using T-coffee.

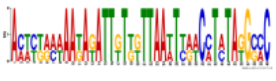
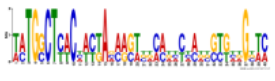
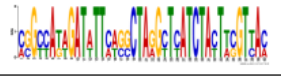
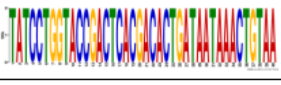
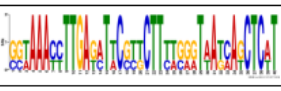
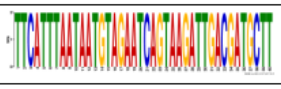
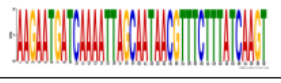
Motif #	E Value	Width (Sites)	Motif Logo	Aptamer Name	Aptamer Sequence	p value
1	4.9E-025	40 (5)		+3-13	CGAACTATGCTGAGGCCCGGCAGATAATGATTACACCAAC	2.68E-24
				+3-24	CGAACTATGCTGAGGCCCGGCAGATAATGATTACACCAAC	2.68E-24
				+3-26	CGAACTATGCTGAGGCCCGGCAGATAATGATTACACCAAC	2.68E-24
				-1-23	TGCAACAAGAAACGCGCAAGGATAGTTAGGATACTGCAAC	1.79E-17
				-1-43	TGCAACAAGAAACGCGCAAGGATAGTTAGGATACTGCAAC	1.79E-17
2	8.1E-019	38 (5)		-1-25	GGACTTTGGGAATATCATGCGGGGTTTGGATAGATGATAA	1.96E-21
				-1-09	GGACTTTGGGAATATCATGCGGGGTTTGGATAGATGATAA	1.96E-21
				-1-04	GGACTTTGGGAATATCATGCGGGGTTTGGATAGATGATAA	1.96E-21
				+2-10	GGGGCAAGTGGATATGATGAAAGGTGGATCTGGGGGGATT	6.38E-17
				+2-08	GGGGCAAGTGGATATGATGAAAGGTGGATCTGGGGGGATT	6.38E-17
3	1.1E-020	40 (5)		-1-17	ACTCTAAAAATAGATTGTTGTTAATTTAACCTATAGGCCC	1.75E-22
				-1-06	ACTCTAAAAATAGATTGTTGTTAATTTAACCTATAGGCCC	1.75E-22
				-1-05	ACTCTAAAAATAGATTGTTGTTAATTTAACCTATAGGCCC	1.75E-22
				+3-07	AAATGGCTAAGATATTTTGTTTAAATCGTCATCTATGGAC	3.46E-17
				+3-01	AAATGGCTAAGATATTTTGTTTAAATCGTCATCTATGGAC	3.46E-17
4	1.2E-013	40 (6)		+1-41B	TATGGCTCTCTACTAACACTGGCAACTCAAGGCGATGCAC	4.79E-15
				+1-04B	TATGGCTCTCTACTAACACTGGCAACTCAAGGCGATGCAC	4.79E-15
				-1-32	AATGCCTTACAACGACAGGACTCAGACACGTCTTCGGGTC	5.11E-14
				-1-22	AATGCCTTACAACGACAGGACTCAGACACGTCTTCGGGTC	5.11E-14
				+3-33	TCTGGCTCACCTTTAGAAGTTAACTTCTATCGTGTAGTTA	7.59E-14
				+3-12	TCTGGCTCACCTTTAGAAGTTAACTTCTATCGTGTAGTTA	7.59E-14
5	1.6E-007	29 (4)		+2-35	TCACGTAAATGGGTATTAGGACCGCATCATCACGGTAAAA	1.87E-15
				+2-25	TCACGTAAATGGGTATTAGGACCGCATCATCACGGTAAAA	1.87E-15
				+2-17	AGACAGAGCACTTAGTTTTCTATTAATAGCGCAATCTCTC	2.50E-14
				+2-05	AGACAGAGCACTTAGTTTTCTATTAATAGCGCAATCTCTC	2.50E-14
6	5.2E-006	40 (3)		+1-56	CGGCCATAGATATTCAGGCTAGGCTCATCTACTTCGTTAC	3.80E-25
				+1-51	CGGCCATAGATATTCAGGCTAGGCTCATCTACTTCGTTAC	3.80E-25
				-3-30	ACGTTAGTGATTTTATCCCTAAGTTTATCTATTAGGTCAA	6.91E-17
7	8.4E-005	40 (2)		+3-48	TATCCTGGTACCGACTCACGACACTGATAATAAACTGTAA	1.20E-24
				+3-02	TATCCTGGTACCGACTCACGACACTGATAATAAACTGTAA	1.20E-24
8	2.8E-004	40 (3)		+3-21	GGTAAACCTTGAGATTCGTTCTTTTGGGTAATCAGCTCAT	2.15E-24
				+3-37	GGTAAACCTTGAGATTCCTTCTTTTGGGTAATCAGCTCAT	1.85E-23
				+3-38	CCAAAATTTTGATCTACGCGCTTCACAATTAGAATCTCTT	1.82E-16
9	1.6E-002	40 (2)		+1-41A	TTCATTTAATAATGTAGAATCAGTAAGATTGACGATGCTT	1.88E-23
				+1-04A	TTCATTTAATAATGTAGAATCAGTAAGATTGACGATGCTT	1.88E-23
10	2.0E-002	40 (2)		+1-55	AAGAATGATCAAAATTAGCAATAACGTTTCTTTATCAAGT	2.12E-23
				+1-10	AAGAATGATCAAAATTAGCAATAACGTTTCTTTATCAAGT	2.12E-23

Figure 5.5.1 List of 10 statistically significant consensus motifs identified via MEME analysis in the central 40-base long segment among CompELS-identified aptamer candidates selected against oxidized MKP3 ($MKP3_{ox}$) and reduced MKP3 ($MKP3_{red}$). For each consensus motif, the E value, the width of the motif, the number of sites the motif is observed, aptamer sequences involved in the motif, the color-coded aptamer sequences, and the P-value of the motif within a particular aptamer sequence is given.

5.5.2 Alignment of MKP3 Sequences

Combining MSA of T-Coffee and the previous MEME analysis allows other distinguishing characteristics of the MKP3 aptamers to be discovered. After alignment aptamer sequences

were color-coded corresponding to their associated identified motif and exact base matches across all sequence members at a specific base position within a motif were underlined.

Table 5.5.1. Multiple sequence alignment of all $MKP3_{ox}$ and $MKP3_{red}$ aptamer candidates using T-Coffee (http://www.ebi.ac.uk/Tools/msa/tcoffee/ accessed June 24, 2017). Base segments of all aptamer sequence candidates that belong to one of the 10 consensus motifs identified separately via MEME analysis are color-coded using the color scheme shown in **Figure 5.5.1**. Members of a motif with exact base matches across all sequence members at a specific position are underlined and any bases not included in a motif (i.e. central segment with a consensus motif less than 40 bases in length) are shown in light gray lettering.

```
+3-01          -AAA-TGGCTAAGATATTTTGTTTAAATC---GTCATCTATGGAC
+3-07          -AAA-TGGCTAAGATATTTTGTTTAAATC---GTCATCTATGGAC
+3-30          -CCG-TGGGCGATGT-TTAAGATGAAGTGG--GATATTTATACTT
+1-37          -GCTGTTACCGA-GCATAAAGTC-AACT--TCGGTATATAGTGAT
+3-25A         -TCAAACTGTAGCTC-CT-TCTGAG-TAAATTCC-TTGCTAGTTC
+2-02          -CGAAATCCTAAGTTAAT-T-TTAGA-GGGAAC-GAGGCCCCATC
+3-43          -CATGTGCTAATC-ATATG-GTG-TAGTATTAG-GACCCAACGTC
+1-38A         CTGGGCGCTTTAA-AGAT-T-TGTGAT-TCCTGTG-CTCATAAAA
+3-02          -TATCCTGG-TACCGACT-CACGACA--CTGATAATAAACTGTAA
+3-48          -TATCCTGG-TACCGACT-CACGACA--CTGATAATAAACTGTAA
+3-23          -TTGCGATGATCAGGGTA-TGAGGCAG-GA-GAAATA-AAGAAAC
+2-11A         -TCGCTTTGTGAA-GGGT-G-CCAGAG-TGGGGGCGGGTTGATGA
-1-04          -GGACTTTGGGAATATCA-TGCGGGGT-TTG--GATAGATGATAA
-1-09          -GGACTTTGGGAATATCA-TGCGGGGT-TTG--GATAGATGATAA
-1-25          -GGACTTTGGGAATATCA-TGCGGGGT-TTG--GATAGATGATAA
+2-11C         -AAATCATAAGAAAGGTTTTGAG-CCTGCGTAT--G-TTCATCAT
+1-06A         -ACAACCTTTC-C-GTTA-TGGGTACCCCAAAGT-CACTCTTTAT
+1-03          -AATTTGTGCGAC-GCTT-TCCGCCAT-GTTAGGA-TTGTCTTGA
+1-38B         -AATGTG-GATAGAGTTA-TAAGCCTC-GTTCG-GGTCATATTAT
+2-20A         -TAAACGTTCTTA-G-TT-TACGTACCTGTACAA-GAGCGATTCC
+1-49          -AAGAAATGCTTAC-CTG-CACGCACC-TT-CATGTATATAGAAT
+3-14          -AAAA-ATAAATGCC-TT-CAACAAGG-ATACTGGTTCACGCTTT
```

Table 5.5.1 continued

-1-23	**-TGCA-ACAAGAAACGCG-CAAGGATA-GT-TAGGATACTGCAAC**
-1-43	**-TGCA-ACAAGAAACGCG-CAAGGATA-GT-TAGGATACTGCAAC**
+2-11B	-AATACTTCA-ACCTATC-TAACGTT-TTCCTGCA-AAAACCATA
+3-08	-AAGAGAC-GTTTAACCTTGGAT-TAAT--AATGAATACAAATTA
+3-19A	-AGAAGTTTGATTA--AA-GAGTCGAA-TGAGGGACTGGTAATTA
+1-06B	-CTTAGACTG-TTTTCTA--GTTAGA-ATGTCTTGCGAGTGCTTA
+2-20B	-CGAGCATCCATTGTGTCTTAAT-GAA---TAGTGCCTTATCTTA
+1-22	-AAAGGATTTAGT-TCTT-AAATTAGCA-TCATT-CAAGATCCAA
+3-20B	-ATCTCATGCAAAGGCATAGGCTC--AAATG--GGATAGTCGAAA
-3-24	-ATCAATTTCAATCATATAC-ACATAATGGTTT---AAGTCTCAA
+3-06	-TGGACTGA-TATAGTGG-AAGGAACCC-G-TTCTTGGAAAACCT
+1-10	**-AAGAATGA-TCAAA-ATTAGCAATAAC-G-TTTCTTTATCAAGT**
+1-55	**-AAGAATGA-TCAAA-ATTAGCAATAAC-G-TTTCTTTATCAAGT**
+3-15	-GAGATTTT-TAAGGTCA-ACTTTTAAT-G-GCATTTTGAGAAAA
+3-28	-GGGCCCAA--AATGTAAGATCTTGGAC-C-AACTTTTAAACAAA
+3-20A	-CGCAAGGT-ATTTTCGA-GTG-GTAA-TTACCTATTTTAATCGC
+2-31	-TGGAT-TTTGAGTG-GA-GGGGATCATTTATTG-GTGTCGTGAC
+3-31	-GGGAT-TGGTATCTGTG-GAGAAAAA-ACG-CTGCTACAATAGA
+2-12	-CAGGAAGTA-CC-CGAA-GTATTAT-AATTTTGAACTACATTCA
+1-25	-AAGAAAAT-GCTGATCT-TGGAAAACT-T-TAATCTTTAAGTCT
-3-12	-TTTAACGT-AGTAA-ACTTTTCAAGTG-T-AAATCTCACAATAT
+2-36	-ACTAGCGCGGTT-GACC-GAAAACGA-TTGTAGA-TTGTGAAAT
-3-30	**-ACGTTAGTGATTTTATC-C-CTAAG-TTTATC-TATTAGGTCAA**
+1-32B	-ACTTTAGA-TTGAGAACAATA-CAGAC-G-CCAATTGACAGGCT
+1-52	-CAATCAGA-GAACG-CCATTTACAGAA-C-TTTAATAAAACAAC
-1-01	-CCC-TAGGTGACACTCCAACCTCAGA---ATTAATGTGTTTAAT
+3-04	-CTTGTTTCCTTA-AAGG-TACA-AGAAGACTGC-ACCCCAGGTT
+3-29A	-CTCG-TATAAGAACACA-T-AAGGT-TGTATCCTACCCTAAGTT
+3-47	-CACGATCGTGAAGACGA-CTC-GATA-GGC-CAAAAGAGAAGTA

Table 5.5.1 continued

```
+3-25B          -AATGGTAC-GCCAT-TTATGGAGGCAA-T-TATACAACGAAATT
+1-30           -AACGTTAGTTTAAGGTT-A-GTCTC-CTTCAGTT-CACCAAATC
-2-02           -TAGCA-GTGATAATGCT-CAGCTCCA-TTAATTA-GGAGAAGTT
+3-16           -GTT-TAAACT-TA--TAACACTGAAGTGAACCCGGGGTGGGATT
+2-18           -GCCTTTA-CGATACGGT-CCTGGTAG-ACTCTTAT-ATGCGATC
+1-17           -ATAGTTCCCTATAA-AT-CCCGGGTC-CTGGGTGT-GTTACTCA
-1-21           -CCCCTT-CAATTTACTT-CACTCCT-AAT-TATGTCGTACCGTA
+3-11           -GACAAGAAGATC-ATCT-CCAGGG-CAAAGTCA-GCCTCNAAAT
-3-29           -GCCAGGTA-AATAACGC-ACA-AGGT-CCACTATTCTCCTATTT
+2-09           -CTCAGTGA-ATACACCG-GT-GCTGGAGTACAA-GCGCGTTAAA
+3-32A          -GA--ATGGTAGTAAGGG-GGACATGT-AGGTCGGGGGCCGAATA
+3-32B          -GGC-ATGACA-CGAGAC-GGCTATCA-GGGGGAGCGGTGGCACC
+2-06           -TAG-ACGAC-ACGATAA-AGGGGTTC-GCCGTAAGGCTGGAACA
+3-19B          --AAATGTAAATTCGACA-CCCGTG-TCGGAGGG-AGGGGTATCA
+2-25           -TC-ACGTAAATGGGTATTAGGA-CCGCATCAT-CA-CGGTAAAA
+2-35           -TC-ACGTAAATGGGTATTAGGA-CCGCATCAT-CA-CGGTAAAA
+1-14B          -TTAAA-AAAAACGTATT-TCCGAGTA-GCA-GTTCACGAGTAGT
+1-32A          -GTAGTGTAA-CGAGA-T-ATAGGGGAGGGGCGG-GAAAGGTATA
+3-21           -GGTAAACC-TTGAGATT-CGT-TCTT-TTGGGTAATCAGCTCAT
+3-37           -GGTAAACC-TTGAGATT-CCT-TCTT-TTGGGTAATCAGCTCAT
+3-29B          -TACTACTCA-TGGTCTT--GCGGCT-GTGGTGGGGGCTGTTTAT
+3-38           -CCAAAATTT-TGATCTA-CGCGCTT-CACAATT-AGAATCTCTT
+1-35           -ATAATAAGGCTGATTT--TGACTTC-AACAGTT-GTCTTGTATT
+1-15           -TACGTTTTTCTTAT-TA-CCCATATAATTGG-T-GTAATGTAAT
+3-27           -GTTGTCAT-TTGAGAGG-GTA-AATC-GATGATGGGTAATTTCT
-3-14           -TGGGTAAA-TCGGTGGA-GTA-GATA-AGTTATCGACGGGTCAT
+1-04A          -TTCATTTA-ATAAT-GTAGAATCAGTA-A-GATTGACGATGCTT
+1-41A          -TTCATTTA-ATAAT-GTAGAATCAGTA-A-GATTGACGATGCTT
-3-13           -TTGTAGTACTTTTA-CG-TTTAGAT--GAGGTTGGGCACCATAA
+3-12           -TCTGGCTCACCTTTAGA-AGTTAAC--TTCTATCGT-GTAGTTA
```

Table 5.5.1 continued

```
+3-33          -TCTGGCTCACCTTTAGA-AGTTAAC--TTCTATCGT-GTAGTTA
+1-04B         -TATGGCTCTCTACTAA--CACTGGCAACTCAA--GGCGATGCAC
+1-41B         -TATGGCTCTCTACTAA--CACTGGCAACTCAA--GGCGATGCAC
+3-41          -TAAACTCCGCTATTAAAGAGATC--GAGTAGG-AA-TTCGTCTT
-1-22          -AATGCCTTACAACGACA-GGA-CTCA-GACAC-GTCTTCGGGTC
-1-32          -AATGCCTTACAACGACA-GGA-CTCA-GACAC-GTCTTCGGGTC
+2-08          -GGGGCAAGTGGATATGA--TGAAAG--GTGGATCTGGGGGGATT
+2-10          -GGGGCAAGTGGATATGA--TGAAAG--GTGGATCTGGGGGGATT
-3-05          -AAAGGCCG-GAGTATGA--TGGATTCAAACCGT-TTCGCTCCTA
+1-26          -ACAGTCTGGC-TCTGGG-TTAGATG--CTAGACCATGATATCTC
+1-40A         -TCTGTTCG-CGTCAAGA-TTAAATTGTATTCGA-T-TGGAGTAC
+1-21          -GATGTCTCCGGCCAGGAGT-GTACTAGTAGAT---AATGGTGAA
-1-18          -GTTTGTTATGACCCGAA-TGTTGGG--CAGAACGA-GTTGGGAT
+3-09          -TCGTAT-ACCGATTCC-GAACTCATATCTCAT--ACTAGCAAAC
+2-15          --GAGCAAGGA-ATTCGTGAGATAACGTCTCAT-CA-TATTAAAC
+1-51          -CGGCCATAGATATTC--AGGCTA--GGCTCATCTACTTCGTTAC
+1-56          -CGGCCATAGATATTC--AGGCTA--GGCTCATCTACTTCGTTAC
+3-18          -GCCACAAGTA-ATATCC-TGCCG-A-ATTATTTAAATTCACAGA
-1-19          -CACTCG-GTCGGTTAC--TGTCGTGCTCTCATC-AACGTATGGC
+3-13          -CGAAC-TATGCTGAGGCCCGGCAGATAATGAT---TACACCAAC
+3-24          -CGAAC-TATGCTGAGGCCCGGCAGATAATGAT---TACACCAAC

+3-26          -CGAAC-TATGCTGAGGCCCGGCAGATAATGAT---TACACCAAC
+1-36          -AGATTAAA-ACTTT-AACAGTCTGAAA-T-AATAACGATCAAAA
+1-14A         -GTAATAAAGG--TAGCT-GGCAAATT-TGGAATATACTTTCATC
+1-40B         -TCAACAGTATCTTAT-TTAGAC-TGTA--GGTTTAACCTTATCC
+3-45          -ATCTAGTCTATGTTAAT-CTAGG-AT-GAAGAT-GTTGTTAAAC
+1-13          -ATGTTGGGGTTGTTAGT-GTAGGGG--AAATTC-CTATTGTGAC
+1-33          -CTTTTTTGT-TTTCAAC-TGA-ATC-AACGAATACTTCTGAGGC
+3-03          -CGAGTGGATCCAC-TTC-T-TTTTG-GTTAGCTTAGTTAGTTTA
+3-10          -ATTTGGAGTGGAG--GT-TATACTTGGTGAGAT-GAATAAGTTA
```

Table 5.5.1 continued

+3-42	-GTTTGA-G-ACTTAGTGATCCCCTT-GCG-TATTATGTACTATT
+3-22	-AGCTGGAA-CT-TTATTGACATCCGAC-A-AATGAGAATTTCTA
+2-05	**-AGACAGAGCACTTAGTT-T--TCTA-TTAATAGCGCAATCTCTC**
+2-17	**-AGACAGAGCACTTAGTT-T--TCTA-TTAATAGCGCAATCTCTC**
+3-17	-TTT-T-AAGATTA-ACCGACTCTGGCTTTTGT-TTCGACCGTTC
+1-01	-CGTATTAAGATTA-AT-GG-TAT-ACTCCAGCCATTAATATGGA
+3-34	-CTT-TATGTACGTGAAT-TATTTAT-ATTGG-GACCAGGGTCAA
+1-29	-CTTTTGTGAAGG-CAGT-A-TTCCAA-TTTTATACGAGGCTTAA
-1-05	**-ACTCTAAAAATA-GATT-GTTGTTAA-TTTAAC-CTATAGGCCC**
-1-06	**-ACTCTAAAAATA-GATT-GTTGTTAA-TTTAAC-CTATAGGCCC**
-1-17	**-ACTCTAAAAATA-GATT-GTTGTTAA-TTTAAC-CTATAGGCCC**
-3-19	-GTATCACTG-TT-GAGG-GATTTATGGCGGACC-CCAGAGGCCC

In this alignment, identical sequences are grouped together for easy identification. While significant conservation of base identities can be observed in subsets of sequences involved in motifs especially in instances with multiple occurrences of the same sequence in a motif, no overarching trends in the alignment between non-"intra-motif" sequences are observed.

5.6 Conclusions

In this Chapter, a two-stage CompELS screening process was used to screen for two conformations of a ROS protein a PTP, MKP3, namely the oxidized and reduced state conformation. 100 aptamers sequences were identified for $MKP3_{ox}$, and 23 aptamer sequences were identified for $MKP3_{red}$. 10 statistically significant motifs were identified with MEME analysis; however, overall alignment between all recovered sequences did not reveal any additional trends. The presented two-stage screening process approach has the potential benefit of a shortened selection process, elimination of intermittent amplification steps, and potential to identify aptamers which can distinguish between even

conformationally related molecular targets as well as identification of individual aptamers for each target in one screening platform.

CHAPTER 6. GENERAL CONCLUSIONS AND FUTURE WORK

6.1 General Conclusions

This dissertation details the development of methods to expand analytical tools for DNA secondary structure analysis of aptamer candidates in conjunction with a non-evolutionary screening process Competition Enhanced Ligand Selection (CompELS). Nucleic acid additions' various effects on the seed mediated growth of gold nanorods (AuNR) is explored to attempt to elucidate interactions of DNA to gold. A non-evolutionary selection process against the AuNR morphology was successful in identifying aptamers sequences, however no clear patterns in primary structure of recovered aptamers emerged. This guided the development of methods to expand the available analytical tools for DNA secondary structure analysis of aptamers. Using a random sequence population as a background for aptamer secondary structure, existing protein sequence alignment tools are adapted here to identify secondary structure domain distributions within aptamers compared to a random sequence background. Building upon previous characterization of secondary structure elements (SSE) in DNA aptamers[293], identified DNA aptamers secondary structures were classified into a 6-point coordinate system and compared against a random sequence population to ascertain significance of an aptamer secondary structure populations. Generation of position specific score matrices (PSSM) enabled scoring of aptamer predicted secondary structures and potential discovery of base position dependent aptamer secondary structure evolution. Lastly, use of a two stage CompELS process for identification of aptamers for reduced and oxidized mitogen-activated protein kinase ($MKP3_{red}$ and $MKP3_{ox}$) targets, demonstrates benefits including a shortened selection

process, elimination of intermittent amplification steps, and potential to identify aptamers which can distinguish between even conformationally related molecular targets as well as identification of individual aptamers for each target in a single screening.

6.2 Significance and Future Work

Methods have been developed and presented in this thesis to expand analysis of groups of DNA aptamer candidates by identifying position-dependent and position-independent secondary structure features and then compare these features to a background generated with a random sequence population. To provide further verification of the validity of these analysis methods, these same tools can be used to analyze other reported aptamers (for the same particular target) with either predicted or experimentally characterized secondary structures. Ideally, such aptamers will already possess previously identified secondary structure motif(s) that serve as the key binding site for the target, such as the sets of different aptamer sequences reported for streptavidin protein from several labs and evaluated by Bing *et al* [47]. Other suggested aptamer sequence sets for validating the current analytical tools include sequence sets identified by Dobbelstein and Shenk[345] or Lozupone *et al.* [346] and later used for validation of structure motif discovery algorithms by Hoinka *et al.* [59] A program called AptaMut to generate individual base substitutions or sequence mutants to identify important features related to binding, such as structural stability, has also been developed by Hoinka *et al.* [347] Thus, this AptaMut program for generation of closely-related variations in the aptamer sequences and their resulting secondary structures would also enable further evaluation and validation of potentially ideal secondary structure motifs. Recently, Dao *et al* published methods [330] related to the presented thesis work for RNA aptamer sequences identified via SELEX and thus could

serve as a parallel analytical tool for evaluating the DNA aptamers identified via a nonevolutionary selection method used in this thesis work. Notably, for validation purposes, aptamer sequences should be of equal total base length and contain the same fixed base region or segment to allow for the generation of an appropriate background sequence library for comparison. Additionally, separate validation focusing on a recurring secondary structure element (e.g. a hairpin secondary structure that was consistently predicted in the fixed base segment of the current work) in which a short, non-base pairing portion of the recurring secondary structure element (e.g. the hairpin loop segment) is randomized could also demonstrate the ability to elicit motifs from ideal structures that are not significantly altered by small changes in the primary structure. An alternate, but related approach for this type of validation in sets of RNA-derived secondary structures from the Rfam database [348] was reported by Hiller *et al.*; however, notably their report exclusively focused on identifying primary structure motifs within these secondary structure elements rather than focusing patterns and motifs on secondary structure elements themselves (independent of base segment identities). [349]

The presented analysis methods have the capability for modification and implementation into high throughput methods of identification of best aptamer candidates, similar to RNA secondary structure analysis suites, such as AptaMotif and APTANI [59, 181] as well as for cases in which few aptamer candidates sequences are recovered. Adaptation of existing protein alignment tools,[170, 333], shows potential for identifying unique structural domains and their distribution in aptamers with the use of generated random sequence populations. Generation of proper background models for secondary structure of aptamer sequences, will also allow for even more advanced identification of secondary structure

domains if combined with analysis suites such as MEME[300, 301] in which expectation maximization (EM) motif searching could be utilized with secondary structure strings (SS$) to further identify secondary structure motifs in aptamers. Additionally, secondary structure analysis in combination with non-evolutionary selection of aptamers can potentially enable creation of screening libraries with particular secondary structure display particular to a given aptamer target after a single screening for next generation aptamer selection.

Future work with identified aptamers for gold nanorods (AuNR) as well as for reduced and oxidized mitogen-activated protein kinase ($MKP3_{red}$ and $MKP3_{ox}$) should include binding studies for relation to the previous analyses presented. AuNR aptamers identified could have great utility in facile functionalization of the particle, with the potential affinity to specific crystalline facets of the AuNR, which may include the {110} or {100} planes. Further analysis of secondary structure of MKP3 aptamers is ongoing in collaboration with Dr. Melissa Kemp.

Evaluation of the two stage CompELS process with next generation sequencing (NGS) could provide further valuable insight into primary and secondary structure roles in affinity generation in non-evolutionary selection, as well as for specificity to either the $MKP3_{red}$ and $MKP3_{ox}$ and their associated conformation. NGS would also provide a meaningful comparison if used simultaneous SELEX and CompELS processes to look at differences in both primary structure evolution and affinity and secondary structure evolution and affinity. If the previous analytical methods for DNA aptamer secondary structure were combined with NGS and CompELS, a greater understanding of relations of affinity generation and aptamer structural evolution could be realized. Finally, adapting CompELS

as well as the analytical tools to include 3D secondary structures could provide a better representation of the true self-folded or self-hybridized aptamer structures for any target of interest.

APPENDIX A.

Table A1. UNAFOLD Predicted Gibbs Free Energies of Aptamer Secondary Structures from Normal Random Libraries

Aptamer Sequence	Structure (S)	dG [kcal/mol]	5% dG Error
101	1	-11.17	-0.56
	2	-11.07	
102	1	-4.70	-0.24
105	1	-6.46	-0.32
108	1	-6.48	-0.32
	2	-6.42	
110	1	-4.12	-0.21
111	1	-3.06	-0.15
	2	-3.02	
	3	-3.00	
112	1	-5.23	-0.26
114	1	-8.36	-0.42
115	1	-4.04	-0.20
116	1	-10.54	-0.53
	2	-10.50	
	3	-10.03	
117	1	-8.19	-0.41
118	1	-7.78	-0.39
119	1	-5.69	-0.28
120	1	-2.70	-0.14
121	1	-6.82	-0.34
201	1	-7.07	-0.35
	2	-6.83	
	3	-6.78	
204	1	-4.70	-0.24
	2	-4.62	
	3	-4.57	
207	1	-6.52	-0.33
210	1	-7.31	-0.37
212	1	-5.21	-0.26
214	1	-4.89	-0.24
217	1	-7.52	-0.38
	2	-7.25	
219	1	-3.85	-0.19

Table A2: UNAFOLD Predicted Gibbs Free Energies of Aptamer Secondary Structures from A-Rich Random Libraries

Aptamer Sequence	Structure (S)	dG [kcal/mol]	5% dG Error
401	1	-6.77	-0.34
402	1	-8.03	-0.40
403	1	-5.08	-0.25
	2	-5.08	
404	1	-6.41	-0.32
405	1	-7.01	-0.35
406	1	-4.72	-0.24
409	1	-7.04	-0.35
410	1	-5.85	-0.29
411	1	-9.08	-0.45
412	1	-3.48	-0.17
	2	-3.31	
413	1	-4.83	-0.24
414	1	-6.89	-0.34
415	1	-5.11	-0.26
	2	-5.04	
416	1	-6.90	-0.35
417	1	-4.61	-0.23
418	1	-7.30	-0.37
419	1	-10.90	-0.55
420	1	-7.31	-0.37
421	1	-4.16	-0.21

REFERENCES

1. Pray, L., *Discovery of DNA structure and function: Watson and Crick.* Nature Education, 2008. **1**(1): p. 100.

2. Alberts, B., et al., *Molecular Biology of the Cell*. 2007: Garland Science.

3. Gmeiner, W.H. and B.J. Walberer, *Base Pairing in DNA: Unusual Patterns*, in *Encyclopedia of Life Sciences*. 2001, John Wiley & Sons, Ltd.

4. Patel, D.J., et al., *Structure, dynamics, and energetics of deoxyguanosine-thymidine wobble base pair formation in the self-complementary d(CGTGAATTCGCG) duplex in solution.* Biochemistry, 2002. **21**(3): p. 437-444.

5. Guest, C.R., et al., *Dynamics of mismatched base pairs in DNA.* Biochemistry, 1991. **30**(13): p. 3271-9.

6. Knauert, M.P. and P.M. Glazer, *Triplex forming oligonucleotides: sequence-specific tools for gene targeting.* Hum Mol Genet, 2001. **10**(20): p. 2243-51.

7. Kaushik, M., et al., *A bouquet of DNA structures: Emerging diversity*. Biochemistry and Biophysics Reports, 2016. **5**: p. 388-395.

8. Burge, S., et al., *Quadruplex DNA: sequence, topology and structure.* Nucleic Acids Res, 2006. **34**(19): p. 5402-15.

9. Lindahl, T. and B. Nyberg, *Rate of depurination of native deoxyribonucleic acid.* Biochemistry, 1972. **11**(19): p. 3610-8.

10. Lett, J.T., E.S. Klucis, and C. Sun, *On the size of the DNA in the mammalian chromosome. Structural subunits.* Biophys J, 1970. **10**(3): p. 277-92.

11. SantaLucia Jr, J. and D. Hicks, *The thermodynamics of DNA structural motifs.* Annu. Rev. Biophys. Biomol. Struct., 2004. **33**: p. 415-440.

12. Wang, G. and K.M. Vasquez, *Impact of alternative DNA structures on DNA damage, DNA repair, and genetic instability.* DNA Repair (Amst), 2014. **19**: p. 143-51.

13. Jin, R., et al., *What controls the melting properties of DNA-linked gold nanoparticle assemblies?* J Am Chem Soc, 2003. **125**(6): p. 1643-54.

14. Park, S.Y. and D. Stroud, *Theory of melting and the optical properties of gold/DNA nanocomposites.* Physical Review B, 2003. **67**(21): p. 212202.

15. Markham, N.R. and M. Zuker, *UNAFold: software for nucleic acid folding and hybridization.* Methods Mol Biol, 2008. **453**: p. 3-31.

16. Dimitrov, R.A. and M. Zuker, *Prediction of hybridization and melting for double-stranded nucleic acids.* Biophys J, 2004. **87**(1): p. 215-26.

17. Rouillard, J.M., M. Zuker, and E. Gulari, *OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach.* Nucleic acids research, 2003. **31**(12): p. 3057-3062.

18. Zuker, M., *Mfold web server for nucleic acid folding and hybridization prediction.* Nucleic Acids Res, 2003. **31**(13): p. 3406-15.

19. Markham, N.R. and M. Zuker, *DINAMelt web server for nucleic acid melting prediction.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W577-81.

20. Zuker, M., *Calculating nucleic acid secondary structure.* Curr Opin Struct Biol, 2000. **10**(3): p. 303-10.

21. Zuker, M., D.H. Mathews, and D.H. Turner, *Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide*, in *RNA biochemistry and biotechnology*. 1999, Springer. p. 11-43.

22. SantaLucia, J., Jr., *A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics.* Proc Natl Acad Sci U S A, 1998. **95**(4): p. 1460-5.

23. Gilbert, S.D., et al., *Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain.* J Mol Biol, 2006. **359**(3): p. 754-68.

24. Gouda, H., et al., *Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods.* Biopolymers, 2003. **68**(1): p. 16-34.

25. Smirnov, I. and R.H. Shafer, *Effect of loop sequence and size on DNA aptamer stability.* Biochemistry, 2000. **39**(6): p. 1462-8.

26. Neves, M.A., et al., *Defining the secondary structural requirements of a cocaine-binding aptamer by a thermodynamic and mutation study.* Biophys Chem, 2010. **153**(1): p. 9-16.

27. Adessi, C., et al., *Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms.* Nucleic Acids Res, 2000. **28**(20): p. E87.

28. Keefe, A.D. and S.T. Cload, *SELEX with modified nucleotides.* Curr Opin Chem Biol, 2008. **12**(4): p. 448-56.

29. Houseley, J. and D. Tollervey, *The many pathways of RNA degradation.* Cell, 2009. **136**(4): p. 763-76.

30. Mikkola, S., et al., *The mechanism of the metal ion promoted cleavage of RNA phosphodiester bonds involves a general acid catalysis by the metal aquo ion on the departure of the leaving group.* Journal of the Chemical Society, Perkin Transactions 2, 1999(8): p. 1619-1626.

31. Vester, B. and J. Wengel, *LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA.* Biochemistry, 2004. **43**(42): p. 13233-41.

32. Kurreck, J., et al., *Design of antisense oligonucleotides stabilized by locked nucleic acids.* Nucleic Acids Res, 2002. **30**(9): p. 1911-8.

33. Petersen, M., et al., *The conformations of locked nucleic acids (LNA).* J Mol Recognit, 2000. **13**(1): p. 44-53.

34. Frieden, M., H.F. Hansen, and T. Koch, *Nuclease stability of LNA oligonucleotides and LNA-DNA chimeras.* Nucleosides Nucleotides Nucleic Acids, 2003. **22**(5-8): p. 1041-3.

35. Eze, N.A. and V.T. Milam, *Exploring locked nucleic acids as a bio-inspired materials assembly and disassembly tool.* Soft Matter, 2013. **9**(8): p. 2403-2411.

36. Eze, N.A., R.S. Sullivan, and V.T. Milam, *Analysis of in Situ LNA and DNA Hybridization Events on Microspheres.* Biomacromolecules, 2017. **18**(4): p. 1086-1096.

37. Ciesiolka, J., J. Gorski, and M. Yarus, *Selection of an RNA domain that binds Zn2+.* RNA, 1995. **1**(5): p. 538-50.

38. Hofmann, H.P., et al., *Ni2+-binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair.* RNA, 1997. **3**(11): p. 1289-300.

39. Zimmermann, G.R., et al., *Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer.* RNA, 2000. **6**(5): p. 659-67.

40. Jo, M., et al., *Development of single-stranded DNA aptamers for specific Bisphenol a detection.* Oligonucleotides, 2011. **21**(2): p. 85-91.

41. Mallikaratchy, P.R., et al., *A multivalent DNA aptamer specific for the B-cell receptor on human lymphoma and leukemia.* Nucleic Acids Res, 2011. **39**(6): p. 2458-69.

42. Liu, J., et al., *Recent developments in protein and cell-targeted aptamer selection and applications.* Curr Med Chem, 2011. **18**(27): p. 4117-25.

43. Radom, F., et al., *Aptamers: molecules of great potential.* Biotechnol Adv, 2013. **31**(8): p. 1260-74.

44. Mann, D., et al., *In vitro selection of DNA aptamers binding ethanolamine.* Biochem Biophys Res Commun, 2005. **338**(4): p. 1928-34.

45. Stoltenburg, R., C. Reinemann, and B. Strehlitz, *SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands.* Biomol Eng, 2007. **24**(4): p. 381-403.

46. Jayasena, S.D., *Aptamers: an emerging class of molecules that rival antibodies in diagnostics.* Clin Chem, 1999. **45**(9): p. 1628-50.

47. Bing, T., et al., *Conservative secondary structure motif of streptavidin-binding aptamers generated by different laboratories.* Bioorg Med Chem, 2010. **18**(5): p. 1798-805.

48. Syed, M.A. and S. Pervaiz, *Advances in aptamers.* Oligonucleotides, 2010. **20**(5): p. 215-24.

49. Hermann, T. and D.J. Patel, *Adaptive recognition by nucleic acid aptamers.* Science, 2000. **287**(5454): p. 820-5.

50. Banerjee, J. and M. Nilsen-Hamilton, *Aptamers: multifunctional molecules for biomedical research.* J Mol Med (Berl), 2013. **91**(12): p. 1333-42.

51. Yang, Y., et al., *Advances in SELEX and application of aptamers in the central nervous system.* Biomol Eng, 2007. **24**(6): p. 583-92.

52. Robertson, D.L. and G.F. Joyce, *Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA.* Nature, 1990. **344**(6265): p. 467-468.

53. Ellington, A.D. and J.W. Szostak, *In vitro selection of RNA molecules that bind specific ligands.* nature, 1990. **346**(6287): p. 818.

54. Tuerk, C. and L. Gold, *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase.* Science, 1990. **249**(4968): p. 505-10.

55. Hamula, C., et al., *Selection and analytical applications of aptamers.* TrAC Trends in Analytical Chemistry, 2006. **25**(7): p. 681-691.

56. Charlton, J. and D. Smith, *Estimation of SELEX pool size by measurement of DNA renaturation rates.* RNA, 1999. **5**(10): p. 1326-32.

57. Schutze, T., et al., *Probing the SELEX process with next-generation sequencing.* PLoS One, 2011. **6**(12): p. e29604.

58. Beier, R., E. Boschke, and D. Labudde, *New strategies for evaluation and analysis of SELEX experiments.* Biomed Res Int, 2014. **2014**: p. 849743.

59. Hoinka, J., et al., *Identification of sequence-structure RNA binding motifs for SELEX-derived aptamers.* Bioinformatics, 2012. **28**(12): p. i215-23.

60. Ouellet, E., et al., *A simple method for eliminating fixed-region interference of aptamer binding during SELEX.* Biotechnol Bioeng, 2014. **111**(11): p. 2265-79.

61. Wilson, D.S. and J.W. Szostak, *In vitro selection of functional nucleic acids.* Annu Rev Biochem, 1999. **68**: p. 611-47.

62. Rimmele, M., *Nucleic acid aptamers as tools and drugs: recent developments.* Chembiochem, 2003. **4**(10): p. 963-71.

63. Guthrie, J.W., et al., *Assays for cytokines using aptamers.* Methods, 2006. **38**(4): p. 324-30.

64. Eulberg, D., et al., *Development of an automated in vitro selection protocol to obtain RNA-based aptamers: identification of a biostable substance P antagonist.* Nucleic Acids Res, 2005. **33**(4): p. e45.

65. Hunniger, T., et al., *Just in time-selection: A rapid semiautomated SELEX of DNA aptamers using magnetic separation and BEAMing.* Anal Chem, 2014. **86**(21): p. 10940-7.

66. Musheev, M.U. and S.N. Krylov, *Selection of aptamers by systematic evolution of ligands by exponential enrichment: addressing the polymerase chain reaction issue.* Anal Chim Acta, 2006. **564**(1): p. 91-6.

67. Thiel, W.H., et al., *Nucleotide bias observed with a short SELEX RNA aptamer library.* Nucleic Acid Ther, 2011. **21**(4): p. 253-63.

68. Tolle, F., et al., *By-product formation in repetitive PCR amplification of DNA libraries during SELEX.* PLoS One, 2014. **9**(12): p. e114693.

69. Yufa, R., et al., *Emulsion PCR significantly improves nonequilibrium capillary electrophoresis of equilibrium mixtures-based aptamer selection: allowing for efficient and rapid selection of aptamer to unmodified ABH2 protein.* Analytical chemistry, 2014. **87**(2): p. 1411-1419.

70. Takahashi, M., et al., *High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency.* Sci Rep, 2016. **6**: p. 33697.

71. Shao, K., et al., *Emulsion PCR: a high efficient way of PCR amplification of random DNA libraries in aptamer selection.* PLoS One, 2011. **6**(9): p. e24910.

72. Xiang, D., et al., *Aptamer-Mediated Cancer Gene Therapy.* Current Gene Therapy, 2015. **15**(2): p. 109-119.

73. Smuc, T., I.Y. Ahn, and H. Ulrich, *Nucleic acid aptamers as high affinity ligands in biotechnology and biosensorics.* J Pharm Biomed Anal, 2013. **81-82**: p. 210-7.

74. Bouchard, P.R., R.M. Hutabarat, and K.M. Thompson, *Discovery and development of therapeutic aptamers.* Annu Rev Pharmacol Toxicol, 2010. **50**: p. 237-57.

75. Ng, E.W., et al., *Pegaptanib, a targeted anti-VEGF aptamer for ocular vascular disease.* Nat Rev Drug Discov, 2006. **5**(2): p. 123-32.

76. Ferrara, N., *Vascular endothelial growth factor: basic science and clinical progress.* Endocr Rev, 2004. **25**(4): p. 581-611.

77. Ferrara, N. and T. Davis-Smyth, *The biology of vascular endothelial growth factor.* Endocr Rev, 1997. **18**(1): p. 4-25.

78. Houck, K.A., et al., *Dual regulation of vascular endothelial growth factor bioavailability by genetic and proteolytic mechanisms.* J Biol Chem, 1992. **267**(36): p. 26031-7.

79. Park, J.E., G.-A. Keller, and N. Ferrara, *The vascular endothelial growth factor (VEGF) isoforms: differential deposition into the subepithelial extracellular matrix and bioactivity of extracellular matrix-bound VEGF.* Molecular biology of the cell, 1993. **4**(12): p. 1317-1326.

80. Wiesmann, C., et al., *Crystal Structure at 1.7 Å Resolution of VEGF in Complex with Domain 2 of the Flt-1 Receptor.* Cell, 1997. **91**(5): p. 695-704.

81. Dunaway, A.B., et al., *Evaluating the dual target binding capabilities of immobilized aptamers using flow cytometry.* Biointerphases, 2015. **10**(1): p. 019015.

82. Keyt, B.A., et al., *The carboxyl-terminal domain (111-165) of vascular endothelial growth factor is critical for its mitogenic potency.* J Biol Chem, 1996. **271**(13): p. 7788-95.

83. Ferrara, N. and W.J. Henzel, *Pituitary follicular cells secrete a novel heparin-binding growth factor specific for vascular endothelial cells.* Biochem Biophys Res Commun, 1989. **161**(2): p. 851-8.

84. Leung, D.W., et al., *Vascular endothelial growth factor is a secreted angiogenic mitogen.* Science, 1989. **246**(4935): p. 1306-9.

85. Connolly, D.T., et al., *Tumor vascular permeability factor stimulates endothelial cell growth and angiogenesis.* J Clin Invest, 1989. **84**(5): p. 1470-8.

86. Keck, P.J., et al., *Vascular permeability factor, an endothelial cell mitogen related to PDGF.* Science, 1989. **246**(4935): p. 1309-12.

87. Plouet, J., J. Schilling, and D. Gospodarowicz, *Isolation and characterization of a newly identified endothelial cell mitogen produced by AtT-20 cells.* EMBO J, 1989. **8**(12): p. 3801-6.

88. Conn, G., et al., *Amino acid and cDNA sequences of a vascular endothelial cell mitogen that is homologous to platelet-derived growth factor.* Proc Natl Acad Sci U S A, 1990. **87**(7): p. 2628-32.

89. Pepper, M.S., et al., *In vitro angiogenic and proteolytic properties of bovine lymphatic endothelial cells.* Exp Cell Res, 1994. **210**(2): p. 298-305.

90. Siddiqui, M.A. and G.M. Keating, *Pegaptanib: in exudative age-related macular degeneration.* Drugs, 2005. **65**(11): p. 1571-7; discussion 1578-9.

91. Ferrara, N., H.P. Gerber, and J. LeCouter, *The biology of VEGF and its receptors.* Nat Med, 2003. **9**(6): p. 669-76.

92. Ferrara, N., et al., *Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene.* Nature, 1996. **380**(6573): p. 439-42.

93. Carmeliet, P., et al., *Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele.* Nature, 1996. **380**(6573): p. 435-9.

94. Gerber, H.P., et al., *VEGF is required for growth and survival in neonatal mice.* Development, 1999. **126**(6): p. 1149-59.

95. Eremina, V., et al., *Glomerular-specific alterations of VEGF-A expression lead to distinct congenital and acquired renal diseases.* J Clin Invest, 2003. **111**(5): p. 707-16.

96. Gerber, H.P., et al., *VEGF couples hypertrophic cartilage remodeling, ossification and angiogenesis during endochondral bone formation.* Nat Med, 1999. **5**(6): p. 623-8.

97. Ryan, A.M., et al., *Preclinical safety evaluation of rhuMAbVEGF, an antiangiogenic humanized monoclonal antibody.* Toxicol Pathol, 1999. **27**(1): p. 78-86.

98. Zimmermann, R.C., et al., *Short-term administration of antivascular endothelial growth factor antibody in the late follicular phase delays follicular development in the rhesus monkey.* J Clin Endocrinol Metab, 2001. **86**(2): p. 768-72.

99. Kim, K.J., et al., *Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumour growth in vivo.* Nature, 1993. **362**(6423): p. 841-4.

100. Klement, G., et al., *Continuous low-dose therapy with vinblastine and VEGF receptor-2 antibody induces sustained tumor regression without overt toxicity.* J Clin Invest, 2000. **105**(8): p. R15-24.

101. Aiello, L.P., et al., *Suppression of retinal neovascularization in vivo by inhibition of vascular endothelial growth factor (VEGF) using soluble VEGF-receptor chimeric proteins.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(23): p. 10457-10461.

102. Adamis, A.P., et al., *Inhibition of vascular endothelial growth factor prevents retinal ischemia-associated iris neovascularization in a nonhuman primate.* Arch Ophthalmol, 1996. **114**(1): p. 66-71.

103. Heier, J.S., et al., *Intravitreal aflibercept (VEGF trap-eye) in wet age-related macular degeneration.* Ophthalmology, 2012. **119**(12): p. 2537-48.

104. Detmar, M., et al., *Increased microvascular density and enhanced leukocyte rolling and adhesion in the skin of VEGF transgenic mice.* J Invest Dermatol, 1998. **111**(1): p. 1-6.

105. Kovacs, Z., et al., *VEGF and flt. Expression time kinetics in rat brain infarct.* Stroke, 1996. **27**(10): p. 1865-72; discussion 1872-3.

106. Ferrara, N., et al., *Differential expression of the angiogenic factor genes vascular endothelial growth factor (VEGF) and endocrine gland-derived VEGF in normal and polycystic human ovaries.* Am J Pathol, 2003. **162**(6): p. 1881-93.

107. Maynard, S.E., et al., *Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension, and proteinuria in preeclampsia.* Journal of Clinical Investigation, 2003. **111**(5): p. 649-658.

108. Moreira, I.S., P.A. Fernandes, and M.J. Ramos, *Vascular endothelial growth factor (VEGF) inhibition--a critical review.* Anticancer Agents Med Chem, 2007. **7**(2): p. 223-45.

109. Giacca, M. and S. Zacchigna, *VEGF gene therapy: therapeutic angiogenesis in the clinic and beyond.* Gene Ther, 2012. **19**(6): p. 622-9.

110. Simon-Yarza, T., et al., *Vascular endothelial growth factor-delivery systems for cardiac repair: an overview.* Theranostics, 2012. **2**(6): p. 541-52.

111. Jiang, F., et al., *Progress and Challenges in Developing Aptamer-Functionalized Targeted Drug Delivery Systems.* Int J Mol Sci, 2015. **16**(10): p. 23784-822.

112. Tan, W., et al., *Molecular aptamers for drug delivery.* Trends Biotechnol, 2011. **29**(12): p. 634-40.

113. Ray, P. and R.R. White, *Aptamers for Targeted Drug Delivery.* Pharmaceuticals (Basel), 2010. **3**(6): p. 1761-1778.

114. Jellinek, D., et al., *Inhibition of receptor binding by high-affinity RNA ligands to vascular endothelial growth factor.* Biochemistry, 1994. **33**(34): p. 10450-6.

115. Green, L.S., et al., *Nuclease-resistant nucleic acid ligands to vascular permeability factor/vascular endothelial growth factor.* Chem Biol, 1995. **2**(10): p. 683-95.

116. Gold, L. and N. Janjic, *High-affinity oligonucleotide ligands to vascular endothelial growth factor (VEGF).* 1998, Google Patents.

117. Ruckman, J., et al., *2'-Fluoropyrimidine RNA-based Aptamers to the 165-Amino Acid Form of Vascular Endothelial Growth Factor (VEGF165).* Journal of Biological Chemistry, 1998. **273**(32): p. 20556-20567.

118. Hasegawa, H., K. Sode, and K. Ikebukuro, *Selection of DNA aptamers against VEGF165 using a protein competitor and the aptamer blotting method.* Biotechnol Lett, 2008. **30**(5): p. 829-34.

119. Potty, A.S., et al., *Biophysical characterization of DNA aptamer interactions with vascular endothelial growth factor.* Biopolymers, 2009. **91**(2): p. 145-56.

120. Kaur, H. and L.Y. Yung, *Probing high affinity sequences of DNA aptamer against VEGF165.* PLoS One, 2012. **7**(2): p. e31196.

121. Choi, E.W., L.V. Nayak, and P.J. Bates, *Cancer-selective antiproliferative activity is a general property of some G-rich oligodeoxynucleotides.* Nucleic Acids Res, 2010. **38**(5): p. 1623-35.

122. Tapp, M., et al., *Spectroscopic studies of nucleic acid additions during seed-mediated growth of gold nanoparticles.* J Mater Res, 2015. **30**(5): p. 666-676.

123. Cui, T. and F. Leng, *Specific recognition of AT-rich DNA sequences by the mammalian high mobility group protein AT-hook 2: a SELEX study.* Biochemistry, 2007. **46**(45): p. 13059-66.

124. Pobanz, K. and A. Luptak, *Improving the odds: Influence of starting pools on in vitro selection outcomes.* Methods, 2016. **106**: p. 14-20.

125. Sabeti, P.C., P.J. Unrau, and D.P. Bartel, *Accessing rare activities from random RNA sequences: the importance of the length of molecules in the starting pool.* Chem Biol, 1997. **4**(10): p. 767-74.

126. Velez, T.E., et al., *Systematic evaluation of the dependence of deoxyribozyme catalysis on random region length.* ACS Comb Sci, 2012. **14**(12): p. 680-7.

127. Stich, M., C. Briones, and S.C. Manrubia, *On the structural repertoire of pools of short, random RNA sequences.* J Theor Biol, 2008. **252**(4): p. 750-63.

128. Vater, A., et al., *Short bioactive Spiegelmers to migraine-associated calcitonin gene-related peptide rapidly identified by a novel approach: tailored-SELEX.* Nucleic Acids Res, 2003. **31**(21): p. e130.

129. Wang, C.Y., et al., *Sequence-constructive SELEX: a new strategy for screening DNA aptamer binding to Globo H.* Biochem Biophys Res Commun, 2014. **452**(3): p. 484-9.

130. He, X., et al., *Stepping Library-Based Post-SELEX Strategy Approaching to the Minimized Aptamer in SPR.* Anal Chem, 2017. **89**(12): p. 6559-6566.

131. Mullis, K.B. and F.A. Faloona, *[21] Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction.* Methods in Enzymology, 1987. **155**: p. 335-350.

132. Wu, D.Y., et al., *The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction.* DNA Cell Biol, 1991. **10**(3): p. 233-8.

133. Jarosch, F., K. Buchner, and S. Klussmann, *In vitro selection using a dual RNA library that allows primerless selection.* Nucleic Acids Res, 2006. **34**(12): p. e86.

134. Pan, W. and G.A. Clawson, *The shorter the better: reducing fixed primer regions of oligonucleotide libraries for aptamer selection.* Molecules, 2009. **14**(4): p. 1353-69.

135. Duan, N., et al., *An ssDNA library immobilized SELEX technique for selection of an aptamer against ractopamine.* Anal Chim Acta, 2017. **961**: p. 100-105.

136. Davis, J.H. and J.W. Szostak, *Isolation of high-affinity GTP aptamers from partially structured RNA libraries.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11616-21.

137. Ruff, K.M., T.M. Snyder, and D.R. Liu, *Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure.* J Am Chem Soc, 2010. **132**(27): p. 9453-64.

138. Nutiu, R. and Y. Li, *In vitro selection of structure-switching signaling aptamers.* Angew Chem Int Ed Engl, 2005. **44**(7): p. 1061-5.

139. Appella, D.H., *Non-natural nucleic acids for synthetic biology.* Curr Opin Chem Biol, 2009. **13**(5-6): p. 687-96.

140. Kang, J., et al., *Combinatorial selection of a single stranded DNA thioaptamer targeting TGF-beta1 protein.* Bioorg Med Chem Lett, 2008. **18**(6): p. 1835-9.

141. Jhaveri, S., B. Olwin, and A.D. Ellington, *In vitro selection of phosphorothiolated aptamers.* Bioorg Med Chem Lett, 1998. **8**(17): p. 2285-90.

142. Mann, A.P., et al., *Identification of thioaptamer ligand against E-selectin: potential application for inflamed vasculature targeting.* PLoS One, 2010. **5**(9): p. e13050.

143. King, D.J., et al., *Novel combinatorial selection of phosphorothioate oligonucleotide aptamers.* Biochemistry, 1998. **37**(47): p. 16489-93.

144. Sharaf, M.L., *Exploring the Immunogenicity and Therapeutic Applications of Boranophosphate-modified RNA: siRNA and RNA Aptamers*, in *Department of Chemistry* 2011, Duke University. p. 198.

145. Charlton, J., G.P. Kirschenheuter, and D. Smith, *Highly potent irreversible inhibitors of neutrophil elastase generated by selection from a randomized DNA-valine phosphonate library.* Biochemistry, 1997. **36**(10): p. 3018-26.

146. Charlton, J., J. Sennello, and D. Smith, *In vivo imaging of inflammation using an aptamer inhibitor of human neutrophil elastase.* Chemistry & Biology, 1997. **4**(11): p. 809-816.

147. Monteith, D.K., et al., *Immune stimulation--a class effect of phosphorothioate oligodeoxynucleotides in rodents.* Anticancer Drug Des, 1997. **12**(5): p. 421-32.

148. Lato, S.M., et al., *Boron-containing aptamers to ATP.* Nucleic Acids Res, 2002. **30**(6): p. 1401-7.

149. Morihiro, K., Y. Kasahara, and S. Obika, *Biological applications of xeno nucleic acids.* Mol Biosyst, 2017. **13**(2): p. 235-245.

150. Kasahara, Y., et al., *2',4'-BNA/LNA aptamers: CE-SELEX using a DNA-based library of full-length 2'-O,4'-C-methylene-bridged/linked bicyclic ribonucleotides.* Bioorg Med Chem Lett, 2013. **23**(5): p. 1288-92.

151. Dellafiore, M.A., J.M. Montserrat, and A.M. Iribarren, *Modified Nucleoside Triphosphates for In-vitro Selection Techniques.* Front Chem, 2016. **4**(18): p. 18.

152. Nonaka, Y., et al., *Affinity improvement of a VEGF aptamer by in silico maturation for a sensitive VEGF-detection system.* Anal Chem, 2013. **85**(2): p. 1132-7.

153. Braden, B.C., et al., *Protein motion and lock and key complementarity in antigen-antibody reactions.* Pharm Acta Helv, 1995. **69**(4): p. 225-30.

154. Koshland, D.E., *The Key–Lock Theory and the Induced Fit Theory.* Angewandte Chemie International Edition in English, 1995. **33**(2324): p. 2375-2378.

155. Ellington, A.D. and J.W. Szostak, *Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures.* Nature, 1992. **355**(6363): p. 850.

156. Forster, U., et al., *Conformational dynamics of the tetracycline-binding aptamer.* Nucleic Acids Res, 2012. **40**(4): p. 1807-17.

157. Bosshard, H.R., *Molecular recognition by induced fit: how fit is the concept?* News Physiol Sci, 2001. **16**: p. 171-3.

158. Vogt, A.D., et al., *Essential role of conformational selection in ligand binding.* Biophys Chem, 2014. **186**: p. 13-21.

159. Sullivan, S.M. and T. Holyoak, *Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection.* Proc Natl Acad Sci U S A, 2008. **105**(37): p. 13829-34.

160. Wlodarski, T. and B. Zagrovic, *Conformational selection and induced fit mechanism underlie specificity in noncovalent interactions with ubiquitin.* Proc Natl Acad Sci U S A, 2009. **106**(46): p. 19346-51.

161. Park, J.W., et al., *Acousto-microfluidics for screening of ssDNA aptamer.* Sci Rep, 2016. **6**: p. 27121.

162. Jing, M. and M.T. Bowser, *Methods for measuring aptamer-protein equilibria: a review.* Anal Chim Acta, 2011. **686**(1-2): p. 9-18.

163. Yao, C., et al., *Development of a quartz crystal microbalance biosensor with aptamers as bio-recognition element.* Sensors, 2010. **10**(6): p. 5859-5871.

164. 7/21/2017]; DNA Oligo Order Form]. Available from: https://www.idtdna.com/Site/Order/oligoentry.

165. Heather, J.M. and B. Chain, *The sequence of sequencers: The history of sequencing DNA.* Genomics, 2016. **107**(1): p. 1-8.

166. Mount, D.W., *Bioinformatics: Sequence and genome analysis*. 2001, Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

167. Needleman, S.B. and C.D. Wunsch, *A general method applicable to the search for similarities in the amino acid sequence of two proteins.* J Mol Biol, 1970. **48**(3): p. 443-53.

168. Smith, T.F. and M.S. Waterman, *Identification of common molecular subsequences.* J Mol Biol, 1981. **147**(1): p. 195-7.

169. Higgins, D.G. and P.M. Sharp, *CLUSTAL: a package for performing multiple sequence alignment on a microcomputer.* Gene, 1988. **73**(1): p. 237-44.

170. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.* Nucleic acids research, 1994. **22**(22): p. 4673-4680.

171. Schneider, T.D. and R.M. Stephens, *Sequence logos: a new way to display consensus sequences.* Nucleic Acids Res, 1990. **18**(20): p. 6097-100.

172. Bailey, T.L. and M. Gribskov, *Combining evidence using p-values: application to sequence homology searches.* Bioinformatics, 1998. **14**(1): p. 48-54.

173. Lawrence, C.E. and A.A. Reilly, *An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences.* Proteins, 1990. **7**(1): p. 41-51.

174. Cardon, L.R. and G.D. Stormo, *Expectation maximization algorithm for identifying protein-binding sites with variable lengths from unaligned DNA fragments.* J Mol Biol, 1992. **223**(1): p. 159-70.

175. Bailey, T.L. and C. Elkan, *Fitting a mixture model by expectation maximization to discover motifs in bipolymers.*

176. Stormo, G.D. and Y. Zhao, *Determining the specificity of protein-DNA interactions.* Nat Rev Genet, 2010. **11**(11): p. 751-60.

177. Davies, S.R., et al., *Computational identification and functional validation of regulatory motifs in cartilage-expressed genes.* Genome Res, 2007. **17**(10): p. 1438-47.

178. McGregor, A., et al., *Secondary Structure Mapping of an RNA Ligand That Has High Affinity for the MetJ Repressor Protein and Interference Modification Analysis of the Protein-RNA Complex.* Journal of Biological Chemistry, 1999. **274**(4): p. 2255-2262.

179. Catherine, A.T., et al., *Rational design of a structure-switching DNA aptamer for potassium ions.* FEBS Open Bio, 2014. **4**: p. 788-95.

180. Liu, J., et al., *A method for aligning RNA secondary structures and its application to RNA motif detection.* BMC Bioinformatics, 2005. **6**(1): p. 89.

181. Caroli, J., et al., *APTANI: a computational tool to select aptamers through sequence-structure motif analysis of HT-SELEX data.* Bioinformatics, 2016. **32**(2): p. 161-4.

182. Alam, K.K., J.L. Chang, and D.H. Burke, *FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections.* Mol Ther Nucleic Acids, 2015. **4**: p. e230.

183. Jiang, P., et al., *MPBind: a Meta-motif-based statistical framework and pipeline to Predict Binding potential of SELEX-derived aptamers.* Bioinformatics, 2014. **30**(18): p. 2665-7.

184. Murphy, C.J., et al., *The Many Faces of Gold Nanorods.* The Journal of Physical Chemistry Letters, 2010. **1**(19): p. 2867-2875.

185. Lohse, S.E. and C.J. Murphy, *The Quest for Shape Control: A History of Gold Nanorod Synthesis.* Chemistry of Materials, 2013. **25**(8): p. 1250-1261.

186. Jain, P.K., et al., *Calculated absorption and scattering properties of gold nanoparticles of different size, shape, and composition: applications in biological imaging and biomedicine.* J Phys Chem B, 2006. **110**(14): p. 7238-48.

187. Xia, F., et al., *Colorimetric detection of DNA, small molecules, proteins, and ions using unmodified gold nanoparticles and conjugated polyelectrolytes.* Proc. Natl. Acad. Sci. USA, 2010. **107**(24): p. 10837-10841.

188. Taton, T.A., C.A. Mirkin, and R.L. Letsinger, *Scanometric DNA array detection with nanoparticle probes.* Science, 2000. **289**(5485): p. 1757-60.

189. Jain, P.K., et al., *Noble metals on the nanoscale: optical and photothermal properties and some applications in imaging, sensing, biology, and medicine.* Acc Chem Res, 2008. **41**(12): p. 1578-86.

190. Storhoff, J.J., et al., *One-Pot Colorimetric Differentiation of Polynucleotides with Single Base Imperfections Using Gold Nanoparticle Probes.* Journal of the American Chemical Society, 1998. **120**(9): p. 1959-1964.

191. Mirkin, C., R. Letsinger, and R. Mucic, *A DNA-based method for rationally assembling nanoparticles into macroscopic materials.* Nature, 1996. **382**: p. 607-609.

192. Gupta, M.K., et al., *Surface assembly and plasmonic properties in strongly coupled segmented gold nanorods.* Small, 2013. **9**(17): p. 2979-90.

193. Sandström, P., M. Boncheva, and B. Åkerman, *Nonspecific and Thiol-Specific Binding of DNA to Gold Nanoparticles.* Langmuir, 2003. **19**(18): p. 7537-7543.

194. Balasubramanian, S.K., et al., *Characterization, purification, and stability of gold nanoparticles.* Biomaterials, 2010. **31**(34): p. 9023-30.

195. Giljohann, D.A., et al., *Gold nanoparticles for biology and medicine.* Angew Chem Int Ed Engl, 2010. **49**(19): p. 3280-94.

196. Huang, X., et al., *Cancer Cell Imaging and Photothermal Therapy in the Near-Infrared Region by Using Gold Nanorods.* J. Am. Chem. Soc., 2006. **128**: p. 2115-2120.

197. Bedford, E.E., et al., *Surface plasmon resonance biosensors incorporating gold nanoparticles.* Macromol Biosci, 2012. **12**(6): p. 724-39.

198. Huang, X., S. Neretina, and M.A. El-Sayed, *Gold nanorods: from synthesis and properties to biological and biomedical applications.* Adv Mater, 2009. **21**(48): p. 4880-4910.

199. Nikoobakht, B. and M.A. El-Sayed, *Preparation and Growth Mechanism of Gold Nanorods (NRs) Using Seed-Mediated Growth Method.* Chemistry of Materials, 2003. **15**(10): p. 1957-1962.

200. Hubert, F., et al., *Nanorods versus nanospheres: a bifurcation mechanism revealed by principal component TEM analysis.* Langmuir, 2010. **26**(10): p. 6887-91.

201. Liu, X., et al., *Extinction coefficient of gold nanoparticles with different sizes and different capping ligands.* Colloids Surf B Biointerfaces, 2007. **58**(1): p. 3-7.

202. Goris, B., et al., *Atomic-scale determination of surface facets in gold nanorods.* Nat Mater, 2012. **11**(11): p. 930-935.

203. Bullen, C., et al., *Chemical kinetics of gold nanorod growth in aqueous CTAB solutions.* Crystal Growth & Design 2011. **11**: p. 3375-3380.

204. Ofir, Y., B. Samanta, and V.M. Rotello, *Polymer and biopolymer mediated self-assembly of gold nanoparticles.* Chem Soc Rev, 2008. **37**(9): p. 1814-25.

205. Ghosh, S.K. and T. Pal, *Interparticle coupling effect on the surface plasmon resonance of gold nanoparticles: from theory to applications.* Chem Rev, 2007. **107**(11): p. 4797-862.

206. Rosi, N.L. and C.A. Mirkin, *Nanostructures in biodiagnostics.* Chem Rev, 2005. **105**(4): p. 1547-62.

207. Geerts, N. and E. Eiser, *DNA-functionalized colloids: Physical properties and applications.* Soft Matter, 2010. **6**(19): p. 4647-4660.

208. Mirkin, C.A., et al., *A DNA-based method for rationally assembling nanoparticles into macroscopic materials.* Nature, 1996. **382**: p. 607-609.

209. Shin, J., X. Zhang, and J. Liu, *DNA-functionalized gold nanoparticles in macromolecularly crowded polymer solutions.* J Phys Chem B, 2012. **116**(45): p. 13396-402.

210. Smith, B.D., et al., *Assembly of DNA-Functionalized Gold Nanoparticles with Gaps and Overhangs in Linker DNA.* The Journal of Physical Chemistry C, 2011. **115**(16): p. 7851-7857.

211. Feldheim, D.L. and B.E. Eaton, *Selection of biomolecules capable of mediating the formation of nanocrystals.* ACS Nano, 2007. **1**(3): p. 154-9.

212. Kim, J., et al., *Peptide-mediated shape- and size-tunable synthesis of gold nanostructures.* Acta Biomater, 2010. **6**(7): p. 2681-9.

213. Brown, S., M. Sarikaya, and E. Johnson, *A genetic analysis of crystal growth.* J Mol Biol, 2000. **299**(3): p. 725-35.

214. Kumar, S.A., Y.A. Peter, and J.L. Nadeau, *Facile biosynthesis, separation and conjugation of gold nanoparticles to doxorubicin.* Nanotechnology, 2008. **19**(49): p. 495101.

215. Gardea-Torresdey, J.L., et al., *Formation and growth of Au nanoparticles inside live alfalfa plants.* Nano Letters, 2002. **2**(4): p. 397-401.

216. Radloff, C., et al., *Metal nanoshell assembly on a virus bioscaffold.* Nano Lett, 2005. **5**(6): p. 1187-91.

217. Berti, L. and G.A. Burley, *Nucleic acid and nucleotide-mediated synthesis of inorganic nanoparticles.* Nat Nanotechnol, 2008. **3**(2): p. 81-7.

218. Bigham, S.R. and J.L. Coffer, *The influence of adenine content on the properties of Q-CdS clusters stabilized by polynucleotides.* Colloids Surf. A, 1995. **95**(2): p. 211-219.

219. Zinchenko, A.A., K. Yoshikawa, and D. Baigl, *DNA-Templated Silver Nanorings.* Advanced Materials, 2005. **17**(23): p. 2820-2823.

220. Lin, Y., et al., *DNA-templated silver nanoparticles as a platform for highly sensitive and selective fluorescence turn-on detection of dopamine.* Small, 2011. **7**(11): p. 1557-61.

221. Wang, Z., et al., *Discovery of the DNA "genetic code" for abiological gold nanoparticle morphologies.* Angew Chem Int Ed Engl, 2012. **51**(36): p. 9078-82.

222. Wang, Z., et al., *DNA-mediated control of metal nanoparticle shape: one-pot synthesis and cellular uptake of highly stable and functional gold nanoflowers.* Nano Lett, 2010. **10**(5): p. 1886-91.

223. Wolf, L.K., Y. Gao, and R.M. Georgiadis, *Sequence-dependent DNA immobilization: specific versus nonspecific contributions.* Langmuir, 2004. **20**(8): p. 3357-61.

224. Liu, J., *Adsorption of DNA onto gold nanoparticles and graphene oxide: surface science and applications.* Phys Chem Chem Phys, 2012. **14**(30): p. 10485-96.

225. Kimura-Suda, H., et al., *Base-dependent competitive adsorption of single-stranded DNA on gold.* J Am Chem Soc, 2003. **125**(30): p. 9014-5.

226. Storhoff, J.J., et al., *Sequence-Dependent Stability of DNA-Modified Gold Nanoparticles.* Langmuir, 2002. **18**(17): p. 6666-6670.

227. Li, H., et al., *Assays Based on Differential Adsorption of Single-stranded and Double-stranded DNA on Unfunctionalized Gold Nanoparticles in a Colloidal Suspension.* Plasmonics, 2007. **2**(4): p. 165-171.

228. Stewart, M.E., et al., *Nanostructured plasmonic sensors.* Chem Rev, 2008. **108**(2): p. 494-521.

229. Link, S. and E. Sayed, *Shape and size dependence of radiative, non-radiative and photothermal properties of gold nanocrystals.* International Reviews In Physical Chemistry, 1999. **19**(3): p. 409-453.

230. Eustis, S. and M.A. El-Sayed, *Determination of the aspect ratio statistical distribution of gold nanorods in solution from a theoretical fit of the observed inhomogeneously broadened longitudinal plasmon resonance absorption spectrum.* J. Appl. Phys., 2006. **100**: p. 044324.

231. Eustis, S. and M. El-Sayed, *Aspect ratio dependence of the enhanced fluorescence intensity of gold nanorods: experimental and simulation study.* J Phys Chem B, 2005. **109**(34): p. 16350-6.

232. Sonnichsen, C., et al., *Plasmon resonances in large noble-metal clusters.* New Journal of Physics, 2002. **4**.

233. Imura, K., et al., *Visualization of Localized Intense Optical Fields in Single Gold−Nanoparticle Assemblies and Ultrasensitive Raman Active Sites.* Nano Letters, 2006. **6**(10): p. 2173-2176.

234. Funston, A.M., et al., *Plasmon coupling of gold nanorods at short distances and in different geometries.* Nano Lett, 2009. **9**(4): p. 1651-8.

235. Jain, P.K., S. Eustis, and M.A. El-Sayed, *Plasmon coupling in nanorod assemblies: optical absorption, discrete dipole approximation simulation, and exciton-coupling model.* J Phys Chem B, 2006. **110**(37): p. 18243-53.

236. Mohamed, M.B., et al., *The 'lightning' gold nanorods: fluorescence enhancement of over a million compared to the gold metal.* Chemical Physics Letters, 2000. **317**(6): p. 517-523.

237. Mayer, K.M., et al., *A label-free immunoassay based upon localized surface plasmon resonance of gold nanorods.* ACS Nano, 2008. **2**(4): p. 687-92.

238. Liu, X., et al., *A one-step homogeneous immunoassay for cancer biomarker detection using gold nanoparticle probes coupled with dynamic light scattering.* J Am Chem Soc, 2008. **130**(9): p. 2780-2.

239. Marinakos, S.M., S. Chen, and A. Chilkoti, *Plasmonic detection of a model analyte in serum by a gold nanorod sensor.* Anal Chem, 2007. **79**(14): p. 5278-83.

240. Nusz, G.J., et al., *Label-free plasmonic detection of biomolecular binding by a single gold nanorod.* Anal Chem, 2008. **80**(4): p. 984-9.

241. Li, X., et al., *Fluorescence quenching of quantum dots by gold nanorods and its application to DNA detection.* Applied Physics Letters, 2009. **94**(6): p. 063111-3.

242. Sendroiu, I.E., M.E. Warner, and R.M. Corn, *Fabrication of silica-coated gold nanorods functionalized with DNA for enhanced surface plasmon resonance imaging biosensing applications.* Langmuir, 2009. **25**(19): p. 11282-4.

243. Wang, Z., et al., *One-step functionalized gold nanorods as intracellular probe with improved SERS performance and reduced cytotoxicity.* Biosens Bioelectron, 2010. **26**(1): p. 241-7.

244. Huang, X., et al., *Cancer cells assemble and align gold nanorods conjugated to antibodies to produce highly enhanced, sharp, and polarized surface Raman spectra: a potential cancer diagnostic marker.* Nano Lett, 2007. **7**(6): p. 1591-7.

245. Oyelere, A.K., et al., *Peptide-conjugated gold nanorods for nuclear targeting.* Bioconjug Chem, 2007. **18**(5): p. 1490-7.

246. PARK, et al., *SERS imaging of HER2-overexpressed MCF7 cells using antibody-conjugated gold nanorods*. 2009, Cambridge, ROYAUME-UNI: Royal Society of Chemistry. 6.

247. Wang, C. and J. Irudayaraj, *Gold nanorod probes for the detection of multiple pathogens.* Small, 2008. **4**(12): p. 2204-8.

248. Parab, H.J., et al., *Biosensing, Cytotoxicity, and Cellular Uptake Studies of Surface-Modified Gold Nanorods.* The Journal of Physical Chemistry C, 2009. **113**(18): p. 7574-7578.

249. Li, Z., et al., *RGD-conjugated dendrimer-modified gold nanorods for in vivo tumor targeting and photothermal therapy.* Mol Pharm, 2010. **7**(1): p. 94-104.

250. Nicholas, J.D., et al. *Two-Photon Luminescence Imaging of Cancer Cells Using Molecularly Targeted Gold Nanorods*. 2007. Optical Society of America.

251. Wang, H., et al., *In vitro and in vivo two-photon luminescence imaging of single gold nanorods.* Proc Natl Acad Sci U S A, 2005. **102**(44): p. 15752-6.

252. Nardelli, A., G. Fronzoni, and M. Stener, *Theoretical Study on the X-ray Absorption at the Sulfur K-Edge in Gold Nanoparticles Protected by Thiolates.* The Journal of Physical Chemistry C, 2009. **113**(33): p. 14844-14851.

253. El Khoury, J.M., et al., *Organo-soluble photoresponsive azothiol monolayer-protected gold nanorods.* Chemical Communications, 2009(16): p. 2109-2111.

254. Pissuwan, D., et al., *A golden bullet? Selective targeting of Toxoplasma gondii tachyzoites using antibody-functionalized gold nanorods.* Nano Lett, 2007. **7**(12): p. 3808-12.

255. Pissuwan, D., et al., *Targeted destruction of murine macrophage cells with bioconjugated gold nanorods.* Journal of Nanoparticle Research, 2007. **9**(6): p. 1109-1124.

256. Perezjuste, J., et al., *Gold nanorods: Synthesis, characterization and applications.* Coordination Chemistry Reviews, 2005. **249**(17-18): p. 1870-1901.

257. Petrova, H., M. Hu, and G.V. Hartland, *Photothermal Properties of Gold Nanoparticles.* Zeitschrift für Physikalische Chemie, 2007. **221**(3): p. 361-376.

258. Li, P.-C., et al., *In vivo photoacoustic molecular imaging withsimultaneous multiple selective targeting usingantibody-conjugated gold nanorods.* Opt. Express, 2008. **16**(23): p. 18605-18615.

259. Wijaya, A., et al., *Selective release of multiple DNA oligonucleotides from gold nanorods.* ACS Nano, 2009. **3**(1): p. 80-6.

260. Huang, Y.F., et al., *Selective photothermal therapy for mixed cancer cells using aptamer-conjugated nanorods.* Langmuir, 2008. **24**(20): p. 11860-5.

261. Wang, C., et al., *Gold nanorod/Fe3O4 nanoparticle "nano-pearl-necklaces" for simultaneous targeting, dual-mode imaging, and photothermal ablation of cancer cells.* Angew Chem Int Ed Engl, 2009. **48**(15): p. 2759-63.

262. Foss, C.A., et al., *Optical properties of composite membranes containing arrays of nanoscopic gold cylinders.* The Journal of Physical Chemistry, 1992. **96**(19): p. 7497-7499.

263. van der Zande, B.M.I., et al., *Colloidal Dispersions of Gold Rods: Synthesis and Optical Properties.* Langmuir, 2000. **16**(2): p. 451-458.

264. Orendorff, C.J. and C.J. Murphy, *Quantitation of metal content in the silver-assisted growth of gold nanorods.* J Phys Chem B, 2006. **110**(9): p. 3990-4.

265. Esumi, K., K. Matsuhisa, and K. Torigoe, *Preparation of Rodlike Gold Particles by UV Irradiation Using Cationic Micelles as a Template.* Langmuir, 1995. **11**(9): p. 3285-3287.

266. Kim, F., J.H. Song, and P. Yang, *Photochemical synthesis of gold nanorods.* J Am Chem Soc, 2002. **124**(48): p. 14316-7.

267. Yu, et al., *Gold Nanorods: Electrochemical Synthesis and Optical Properties.* The Journal of Physical Chemistry B, 1997. **101**(34): p. 6661-6664.

268. Liu, M. and P. Guyot-Sionnest, *Mechanism of silver(I)-assisted growth of gold nanorods and bipyramids.* J Phys Chem B, 2005. **109**(47): p. 22192-200.

269. Jana, N.R., L. Gearheart, and C.J. Murphy, *Wet Chemical Synthesis of High Aspect Ratio Cylindrical Gold Nanorods.* The Journal of Physical Chemistry B, 2001. **105**(19): p. 4065-4067.

270. Zijlstra, P., et al., *High-temperature seedless synthesis of gold nanorods.* J Phys Chem B, 2006. **110**(39): p. 19315-8.

271. Jana, N.R., *Gram-scale synthesis of soluble, near-monodisperse gold nanorods and other anisotropic nanoparticles.* Small, 2005. **1**(8-9): p. 875-82.

272. Chandran, S.P., et al., *Synthesis of gold nanotriangles and silver nanoparticles using Aloe vera plant extract.* Biotechnol Prog, 2006. **22**(2): p. 577-83.

273. Canizal, G., et al., *Multiple Twinned Gold Nanorods Grown by Bio-reduction Techniques.* Journal of Nanoparticle Research, 2001. **3**(5/6): p. 475-481.

274. Shankar, S.S., et al., *Biological synthesis of triangular gold nanoprisms.* Nat Mater, 2004. **3**(7): p. 482-8.

275. Murphy, C.J., et al., *Gold nanorod crystal growth: From seed-mediated synthesis to nanoscale sculpting.* Current Opinion in Colloid & Interface Science, 2011. **16**(2): p. 128-134.

276. Lee, K.S. and M.A. El-Sayed, *Gold and Silver Nanoparticles in Sensing and Imaging: Sensitivity of Plasmon Response to Size, Shape, and Metal Composition.* J. Phys. Chem. B, 2006. **110**(39): p. 19220-19225.

277. Zweifel, D.A. and A. Wei, *Sulfide-Arrested Growth of Gold Nanorods.* Chem. Mater., 2005. **17**(16): p. 4256-4261.

278. Rozenblum, G.T., et al., *Aptamers: current challenges and future prospects.* Expert Opin. Drug Discovery, 2016. **11**(2): p. 127-135.

279. Lauhon, C.T. and J.W. Szostak, *RNA aptamers that bind flavin and nicotinamide redox cofactors.* J. Am. Chem. Soc., 1995. **117**(4): p. 1246-1257.

280. Cox, J.C. and A.D. Ellington, *Automated selections of anti-protein aptamers.* Bioorg. Med. Chem., 2001. **9**: p. 2525-2531.

281. Cox, J.C., P. Rudolph, and A.D. Ellington, *Automated RNA selection.* Biotechnol. Prog., 1998. **14**: p. 845-850.

282. Torabi, S.-F., et al., *In vitro selection of a sodium-specific DNAzyme and its application in intracellular sensing.* Proc. Natl. Acad. Sci. USA, 2015. **112**(19): p. 5903-5908.

283. Reinholt, S.J., et al., *Highly multiplexed RNA aptamer delection using a microplate-based microcolumn device.* Sci. Rep., 2016. **6**.

284. Szeto, K., et al., *RAPID-SELEX for RNA aptamers.* PLoS ONE, 2013. **8**(12): p. e82667.

285. Arnold, S., et al., *One round of SELEX for the generation of DNA aptamers directed against KLK6.* Biol. Chem., 2012. **393**: p. 343-353.

286. Lai, J.-C. and C.-Y. Hong, *Magnetic-assisted rapid aptamer selection (MARAS) for generating high-affinity DNA aptamer using rotating magnetic fields.* ACS Combinatorial Science, 2014. **16**: p. 321-327.

287. Boese, B.J., K. Corbino, and R.R. Breaker, *In vitro selection and characterization of cellulose-binding RNA aptamers using isothermal amplification.* Nucleosides Nucleotides Nucleic Acids, 2008. **27**(8): p. 949-66.

288. Tanaka, Y., et al., *RNA aptamers targeting the carboxyl terminus of KRAS oncoprotein generated by an improved SELEX with isothermal RNA amplification.* Oligonucleotides, 2007. **17**(1): p. 12-21.

289. Boiziau, C., et al., *DNA Aptamers Selected Against the HIV-1trans-Activation-responsive RNA Element Form RNA-DNA Kissing Complexes.* Journal of Biological Chemistry, 1999. **274**(18): p. 12730-12737.

290. Bibby, D.F., et al., *Application of a novel in vitro selection technique to isolate and characterise high affinity DNA aptamers binding mammalian prion proteins.* J Virol Methods, 2008. **151**(1): p. 107-15.

291. Stevenson, R., et al., *Binding of 14-3-3 proteins to a single stranded oligodeoxynucleotide aptamer.* Bioorganic chemistry, 2008. **36**(5): p. 215-219.

292. Kim, Y.S., et al., *Isolation and characterization of enantioselective DNA aptamers for ibuprofen.* Bioorg Med Chem, 2010. **18**(10): p. 3467-73.

293. Tapp, M., *Competition-Induced Selection of Ligands for the Screening of DNA Aptamers for Gold Substrates*, in *Materials Science and Engineering*. 2015, Georgia Institute of Technology.

294. Tapp, M.J.N., et al., *Competition-enhanced ligand selection to identify DNA aptamers.* in preparation.

295. Zhang, Q., et al., *Faceted Gold Nanorods: Nanocuboids, Convex Nanocuboids, and Concave Nanocuboids.* Nano Letters, 2015. **15**(6): p. 4161-4169.

296. Katz-Boon, H., et al., *Three-Dimensional Morphology and Crystallography of Gold Nanorods.* Nano Letters, 2011. **11**(1): p. 273-278.

297. Carbó-Argibay, E., et al., *The Crystalline Structure of Gold Nanorods Revisited: Evidence for Higher-Index Lateral Facets.* Angewandte Chemie International Edition, 2010. **49**(49): p. 9397-9400.

298. Gou, L. and C.J. Murphy, *Fine-tuning the shape of gold nanorods.* Chem. Mater., 2005. **17**(14): p. 3668-3672.

299. Yushi, T., et al., *Tuned longitudinal surface plasmon resonance and third-order nonlinear optical properties of gold nanorods.* Nanotechnology, 2011. **22**(27): p. 275203.

300. Bailey, T.L., et al., *MEME SUITE: tools for motif discovery and searching.* Nucleic Acids Res., 2009. **37**(Web Server): p. W202-W208.

301. Bailey, T.L., et al., *MEME: discovering and analyzing DNA and protein sequence motifs.* Nucleic Acids Res., 2006. **34**(Web Server): p. W369-W373.

302. Notredame, C., D.G. Higgins, and J. Heringa, *T-Coffee: A novel method for fast and accurate multiple sequence alignment.* J. Mol. Biol., 2000. **302**: p. 205-217.

303. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res., 2004. **32**: p. 1792–1797.

304. Noorden, R.V., B. Maher, and R. Nuzzo, *The top 100 papers.* Nature, 2014. **514**: p. 550-553.

305. Di Tommaso, P., et al., *T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension.* Nucleic Acids Res., 2011. **39**: p. W13-17.

306. Wheeler, T.J. and J.D. Kececioglu, *Multiple alignment by aligning alignments.* Bioinformatics, 2007. **23**: p. i559–i568.

307. Stoltenburg, R., N. Nikolaus, and B. Strehlitz, *Capture-SELEX: Selection of DNA Aptamers for aminoglycoside antibiotics.* J. Anal. Methods Chemistry, 2012. **2012**: p. 415697.

308. Kelly, J.A., J. Feignon, and T.O. Yeates, *Reconciliation of the x-ray and NMR structures of the thrombin-binding aptamer d(GGTTGGTGTGGTTGG).* J. Mol. Biol., 1996. **256**: p. 417-422.

309. Macaya, R.F., et al., *Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution.* Proc. Natl. Acad. Sci. USA, 1993. **90**(8): p. 3745-3749.

310. Padmanabhan, K., et al., *The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer.* J. Biol. Chem., 1993. **268**(24): p. 17651-17654.

311. Shangguan, D., et al., *Optimization and modifications of aptamers selected from live cancer cell lines.* ChemBioChem, 2007. **8**: p. 603-606.

312. Rockey, W.M., et al., *Rational truncation of an RNA aptamer to prostate-specific membrane antigen using computational structural modeling.* Nucleic Acid Ther., 2011. **21**(5): p. 299-314.

313. Berezhnoy, A., et al., *Isolation and optimization of murine IL-10 receptor blocking oligonucleotide aptamers using high-throughput sequencing.* Mol. Ther., 2012. **20**(6): p. 1242-1250.

314. Le, T.T., O. Chumphukan, and A.E.G. Cass, *Determination of minimal sequence for binding of an aptamer. A comparison of truncation and hybridization inhibition methods.* RSC Adv., 2014. **4**: p. 47227–47233.

315. Klosterman, P.S., et al., *Three-dimensional motifs from the SCOR, structural classification of RNA database: extruded strands, base triples, tetraloops and U-turns.* Nucleic Acids Res., 2004. **32**(8): p. 2342-2352.

316. Bindewald, E., et al., *RNAJunction: a database of RNA junctions and kissing loops for three-dimensional structural analysis and nanodesign.* Nucleic Acids Res., 2007. **36**: p. D392-D397.

317. Shulman-Peleg, A., R. Nussinov, and H.J. Wolfson, *RsiteDB: a database of protein binding pockets that interact with RNA nucleotide bases.* Nucleic Acids Res., 2009. **37**: p. D369-D373.

318. Yi, Y., et al., *RAID v2.0: an updated resource of RNA-associated interactions across organisms.* Nucleic Acids Res., 2017. **45**(D1): p. D115-D118.

319. Jiang, F., et al., *Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex.* Nature, 1996. **382**(6587): p. 183-186.

320. Yang, Y., et al., *Structural basis of ligand discrimination by two related RNA aptamers resolved by NMR spectroscopy.* Science, 1996. **272**(5266): p. 1343-1347.

321. Mathews, D.H. and D.H. Turner, *Dyalign: an algorithm for finding the secondary structure common to two RNA sequences.* Journal of Molecular Biology, 2002. **317**.

322. Mathews, D.H., *Predicting a set of minimal free energy RNA secondary structures common to two sequences.* Bioinformatics, 2005. **21**(10): p. 2246-2253.

323. Carothers, J.M., et al., *Informational complexity and functional activity of RNA structures.* J. Am. Chem. Soc., 2004. **126**: p. 5130-5137.

324. Tabei, Y., et al., *A fast structural multiple alignment method for long RNA sequences.* BMC Bioinformatics, 2008. **9**: p. 33.

325. Will, S., et al., *SPARSE: quadratic time simultaneous alignment and folding of RNAs without sequence-based heuristics.* Bioinformatics, 2015. **31**: p. 2489-2496.

326. Rogers, E. and C. Heitsch, *New insights from cluster analysis methods for RNA secondary structure prediction.* WIREs RNA, 2016. **7**: p. 278-294.

327. Kemena, C., et al., *Using tertiary structure for the computation of highly accurate multiple RNA alignments with the SARA-Coffee package.* Bioinformatics 2013. **29**(9): p. 1112–1119.

328. Sorescu, D.A., et al., *CARNA-alignment of RNA structure ensembles.* Nucl. Acids Res., 2012. **40**(W1): p. W49-W53.

329. Pei, S., B.L. Slinger, and M.M. Meyer, *Recognizing RNA structural motifs in HT-SELEX data for ribosomal protein S15.* BMC Bioinformatics, 2017. **18**: p. 298.

330. Dao, P., et al., *AptaTRACE elucidates RNA sequence-structure motifs from selection trends in HT-SELEX experiments.* Cell Systems, 2016. **3**: p. 62-70.

331. Mélard, G., *On the accuracy of statistical procedures in Microsoft Excel 2010.* Computational statistics, 2014. **29**(5): p. 1095-1128.

332. Farrance, I. and R. Frenkel, *Uncertainty in Measurement: A Review of Monte Carlo Simulation Using Microsoft Excel for the Calculation of Uncertainties Through Functional Relationships, Including Uncertainties in Empirically Derived Constants.* The Clinical Biochemist Reviews, 2014. **35**(1): p. 37-61.

333. Kumar, S., G. Stecher, and K. Tamura, *MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets.* Molecular Biology and Evolution, 2016. **33**(7): p. 1870-1874.

334. Bock, L.C., et al., *Selection of single-stranded DNA molecules that bind and inhibit human thrombin.* Nature, 1992. **355**: p. 564-566.

335. Gragoudas, E.S., et al., *Pegaptanib for neovascular age-related macular degeneration.* N. Engl. J. Med., 2004. **351**(27): p. 2805-2816.

336. Nonaka, Y., K. Sode, and K. Ikebukuro, *Screening and Improvement of an Anti-VEGF DNA Aptamer*. Molecules, 2010. **15**(1): p. 215-225.

337. Padmanabhan, K. and A. Tulinsky, *An ambiguous structure of a DNA 15-mer thrombin complex.* Acta. Cryst., 1996. **D52**: p. 272-282.

338. Lee, J.-H., F. Jucker, and A. Pardi, *Imino proton exchange rates imply an induced-fit binding mechanism for the VEGF165-targeting aptamer, Macugen.* FEBS Letters, 2009. **582**: p. 1835-1839.

339. Muller, Y.A., et al., *Vascular endothelial growth factor: Crystal structure and functional mapping of the kinase domain receptor binding site.* Proc. Natl. Acad. Sci. USA, 1997. **94**(14): p. 7192-7197.

340. Gelinas, A.D., D.R. Davies, and N. Janjic, *Embracing proteins: structural themes in aptamer–protein complexes.* Curr. Opin. Struct. Biol., 2016. **36**: p. 122-132.

341. Lou, Y.W., et al., *Redox regulation of the protein tyrosine phosphatase PTP1B in cancer cells.* FEBS J, 2008. **275**(1): p. 69-88.

342. den Hertog, J., A. Groen, and T. van der Wijk, *Redox regulation of protein-tyrosine phosphatases.* Arch Biochem Biophys, 2005. **434**(1): p. 11-5.

343. Seth, D. and J. Rudolph, *Redox regulation of MAP kinase phosphatase 3.* Biochemistry, 2006. **45**(28): p. 8476-87.

344. Bose, A.K. and K.A. Janes, *A high-throughput assay for phosphoprotein-specific phosphatase activity in cellular extracts.* Mol Cell Proteomics, 2013. **12**(3): p. 797-806.

345. Dobbelstein, M. and T. Shenk, *In vitro selection of RNA ligands for the ribosomal L22 protein associated with Epstein-Barr virus-expressed RNA by using randomized and cDNA-derived RNA libraries.* J Virol, 1995. **69**(12): p. 8027-34.

346. LOZUPONE, C., et al., *Selection of the simplest RNA that binds isoleucine.* RNA, 2003. **9**(11): p. 1315-1322.

347. Hoinka, J., et al., *Large scale analysis of the mutational landscape in HT-SELEX improves aptamer discovery.* Nucleic Acids Research, 2015. **43**(12): p. 5699-5707.

348. Griffiths-Jones, S., et al., *Rfam: annotating non-coding RNAs in complete genomes.* Nucleic Acids Research, 2005. **33**(suppl_1): p. D121-D124.

349. Hiller, M., et al., *Using RNA secondary structures to guide sequence motif finding towards single-stranded regions.* Nucleic Acids Research, 2006. **34**(17): p. e117-e117.